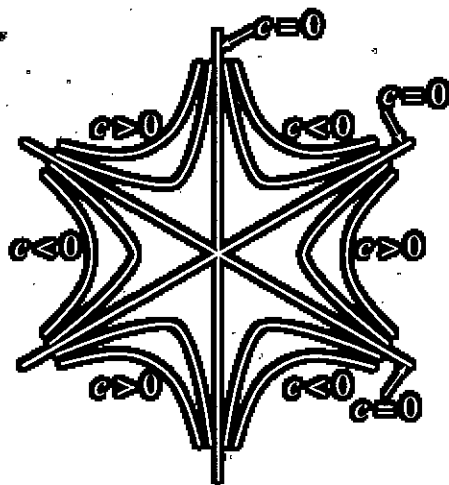


VOLUME 2

CÁLCULO

*Cálculo com funções de várias variáveis
e Álgebra Linear, com aplicações às
equações diferenciais e às probabilidades*

TOM M. APOSTOL



EDITORIAL REVERTÉ

Cálculo

TOM M. APOSTOL

CÁLCULO

VOLUME 2

*Cálculo com funções de várias variáveis
e Álgebra Linear, com aplicações às
equações diferenciais e às probabilidades*



EDITORIAL REVERTÉ, S. A.

Barcelona - Bogotá - Buenos Aires - Caracas - México

Título da obra original:

**CALCULUS, One-Variable Calculus,
With an introduction to Linear Algebra**
Second Edition. Volume 2

Edição original em língua inglesa publicada por:

Blaisdell Publishing Company, Waltham, Massachusetts, USA

Copyright © by Blaisdell Publishing Company

Tradução de:

Joaquim Ferreira Marques

Doutor em Ciências Exactas

Propiedad de:

EDITORIAL REVERTÉ, S. A.

Loreto, 13-15, Local B

08029 Barcelona

Tel: (34) 93 419 33 36

Fax: (34) 93 419 51 89

e-mail: reverte@reverte.com

www.reverte.com

Proibida a reprodução de toda ou parte desta obra, sob qualquer forma, sem autorização por escrito do editor.

Reservados todos os direitos

Edição em português

© EDITORIAL REVERTÉ, S. A., 1996

Reimpresión: octubre de 2004

Impreso en España - Printed in Spain

ISBN: 84-291-5016-1 Torno 2

ISBN: 84-291-5014-5 Obra completa

Depósito Legal: B-44494-2004

Impreso por Domingraf Impressors

Pol. Ind. Can Magarola

08100 Mollet del Vallés (Barcelona)

a

Jane e Stephen

PREFÁCIO

Este livro é a continuação do livro do autor *Cálculo, volume I, Segunda Edição*. O presente volume foi escrito com a mesma ideia fundamental que norteou o primeiro. Uma adequada orientação para a técnica ligada a um rigoroso e profundo desenvolvimento teórico. Procurou-se fazer chegar ao estudante o espírito da matemática moderna sem exagerar o formalismo. Como no Volume I, incluem-se notas históricas para dar ao estudante uma ideia da evolução do pensamento matemático.

O segundo volume está dividido em três partes, intituladas *Análise Linear*, *Análise não Linear* e *Tópicos Especiais*. Os dois últimos capítulos do Volume I repetem-se aqui, constituindo os dois primeiros capítulos deste Volume, com a finalidade de que todo o material relativo à Álgebra Linear se apresenta de forma completa em cada um dos volumes.

A Parte 1 contém uma introdução à álgebra linear, incluindo transformações lineares, matrizes, determinantes, valores próprios e formas quadráticas. Fazem-se algumas aplicações à Análise, em particular ao estudo das equações diferenciais lineares. Com a ajuda do cálculo matricial estudam-se os sistemas de equações diferenciais. Demonstram-se teoremas de existência e unicidade por intermédio do método de Picard das aproximações sucessivas, que também se trata na teoria dos operadores de contracção.

Na Parte 2 estuda-se o cálculo para funções de várias variáveis. O cálculo diferencial é unificado e simplificado com auxílio da álgebra linear. Incluem-se a generalização da regra de derivação de uma função composta para campos vectoriais e escalares e aplicações às equações de derivadas parciais e a problemas de extremos. O cálculo integral inclui os integrais de linha, integrais múltiplos, e integrais de superfície, com aplicações à Análise vectorial. Aqui a exposição segue mais ou menos a linha clássica e não inclui um desenvolvimento formal das formas diferenciais.

Os tópicos especiais tratados na Parte 3 são *Probabilidades e Análise Numérica*. A parte referente às Probabilidades está dividida em dois capítulos, um que trata o

assunto considerando o conjunto fundamental (ou espaço amostra) finito ou infinito numerável; o outro em que se consideram conjuntos fundamentais não numeráveis, variáveis aleatórias e funções de repartição. Fazem-se algumas aplicações no estudo de variáveis aleatórias uni e bidimensionais.

O último capítulo contém uma introdução à Análise Numérica, dando-se particular ênfase ao estudo de diferentes tipos de aproximação polinomial. Aqui, mais uma vez se procura a unificação das ideias pela notação e terminologia da álgebra linear. O livro termina com o estudo de fórmulas de integração aproximada, tais como a regra de Simpson, e com uma discussão da fórmula de somação de Euler.

Contém este volume matéria suficiente para um curso anual com três ou quatro tempos semanais. Pressupõe a conhecimento do cálculo para funções de uma variável tal como se estuda na maior parte dos primeiros anos dos cursos de cálculo. O autor idealizou a matéria exposta para um curso com quatro aulas semanais, duas de exposição por parte do professor e duas para questões postas aos alunos, desenvolvido ao longo de dez semanas para cada parte e omitindo as secções assinaladas com um asterisco.

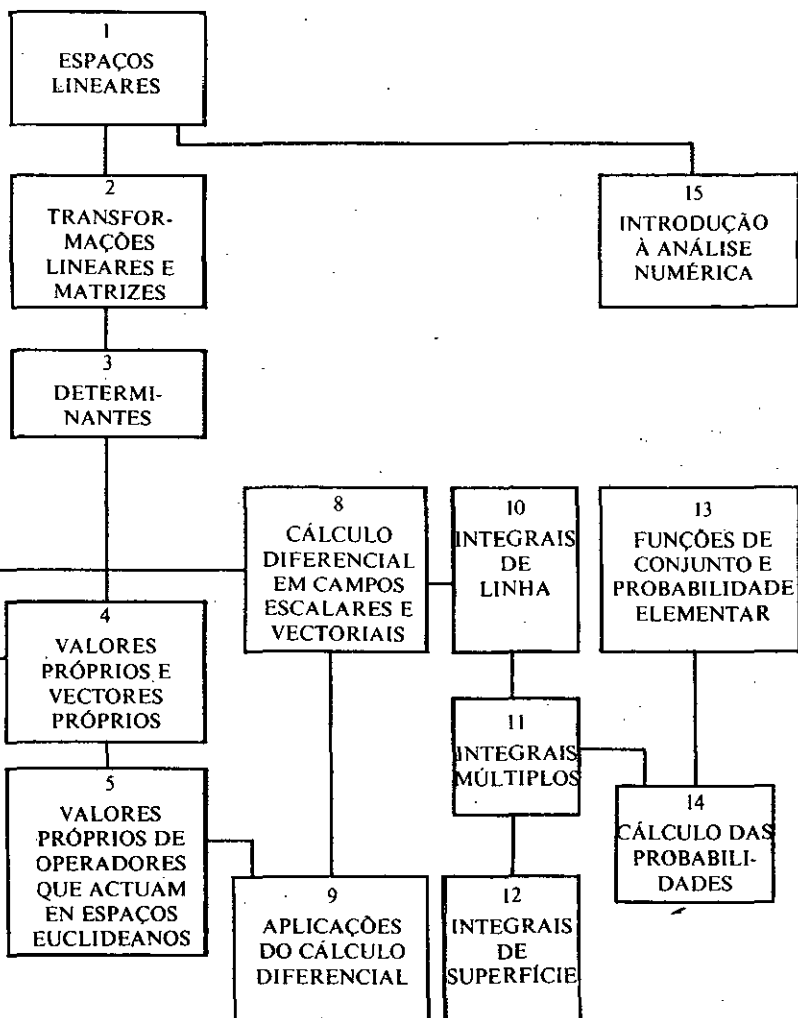
Este segundo volume foi planeado de maneira a poderem omitirse vários capítulos em cursos abreviados. Por exemplo, o último capítulo de cada uma das partes pode ser omitido, sem que tal origine descontinuidade na exposição. A Parte I proporciona material para um curso combinado de álgebra linear e equações diferenciais ordinárias. Cada professor pode escolher os tópicos adequados às suas necessidades e preferências por consulta do diagrama da página seguinte que coindencia a interdependência lógica dos capítulos.

Mais uma vez agradeço com prazer a colaboração de muitos amigos e colegas. Ao preparar a segunda edição recebi valiosa ajuda dos Professores Herbert S. Zuckerman da Universidade de Washington e Basil Gordon da Universidade da Califórnia, Los Angeles, tendo cada um deles sugerido várias modificações. Agradecimento são também devidos ao pessoal da Blaisdell Publishing Company pela sua assistência e cooperação.

Como noutras ocasiões, é para mim uma satisfação especial exprimir a minha gratidão a minha esposa pela sua valiosa e variada colaboração. Em sinal de reconhecimento dedico-lhe gostosamente este livro.

T. M. A.

Pasadena, Califórnia



ÍNDICE ANALÍTICO

PARTE 1. ANÁLISE LINEAR

1. ESPAÇOS LINEARES

1.1.	Introdução	3
1.2.	Definição de espaço linear	3
1.3.	Exemplos de espaços lineares	5
1.4.	Consequências elementares dos axiomas	6
1.5.	Exercícios	8
1.6.	Subespaços de um espaço linear	9
1.7.	Conjuntos dependentes e independentes num espaço linear	10
1.8.	Bases e dimensão	13
1.9.	Componentes	15
1.10.	Exercícios	15
1.11.	Productio interno, espaços euclidianos. Normas	16
1.12.	Ortogonalidade num espaço euclidiano	20
1.13.	Exercícios	23
1.14.	Construção de conjuntos ortogonais. O método de Gram-Schmidt	25
1.15.	Complementos ortogonais. Projecções	30
1.16.	A melhor aproximação de elementos de um espaço euclidiano por elementos de um subespaço de dimensão finita	32
1.17.	Exercícios	34

2. TRANSFORMAÇÕES LINEARES E MATRIZES

2.1.	Transformações lineares	35
2.2.	Espaço nulo e contradomínio	37
2.3.	Nulidade e ordem	38

2.4.	Exercícios	39	
2.5.	Operações algébricas relativas a transformações lineares	41	
2.6.	Inversas	43	
2.7.	Transformações lineares biunívocas	46	
2.8.	Exercícios	48	
2.9.	Transformações lineares com valores determinados	50	
2.10.	Representação matricial das transformações lineares	51	
2.11.	Construção de uma representação matricial na forma diagonal	54	
2.12.	Exercícios	56	
2.13.	Espaços lineares de matrizes	58	
2.14.	Isomorfismo entre transformações lineares de matrizes	59	
2.15.	Multiplicação de matrizes	61	
2.16.	Exercícios	64	
2.17.	Sistemas de equações lineares	66	
2.18.	Técnicas de cálculo	68	
2.19.	Inversas de matrizes quadradas	73	
2.20.	Exercícios	76	
2.21.	Exercícios variados sobre matrizes	77	

3. DETERMINANTES

3.1.	Introdução	79	
3.2.	Justificação da escolha dos axiomas para a função determinante	80	
3.3.	Um conjunto de axiomas para a função determinante	82	
3.4.	Cálculo de determinantes	84	
3.5.	O teorema de unicidade	88	
3.6.	Exercícios	89	
3.7.	Producto de determinantes	91	
3.8.	Determinante da matriz inversa de uma matriz não singular	92	
3.9.	Determinantes e independência de vectores	93	
3.10.	Determinante de uma matriz diagonal por blocos	93	
3.11.	Exercícios	95	
3.12.	Fórmulas para o desenvolvimento de determinantes. Menores e complementos algébricos	96	
3.13.	Existência da função determinante	100	
3.14.	O determinante da matriz transposta	102	
3.15.	A matriz complementos algébricos	103	
3.16.	Regra de Cramer	105	
3.17.	Exercícios	106	

4. VALORES PRÓPRIOS E VECTORES PRÓPRIOS

4.1.	Transformações lineares representadas por matrizes diagonais	109	
4.2.	Valores próprios e vectores próprios de uma transformação linear	110	
4.3.	Independência linear de vectores próprios correspondentes a valores próprios distintos	113	
4.4.	Exercícios	113	
4.5.	O caso de dimensão finita. Polinómios característicos	116	

- 4.6. Cálculo de valores próprios e vectores próprios no caso de dimensão finita 117
- 4.7. Traço de uma matriz 120
- 4.8. Exercícios 121
- 4.9. Matrizes representando a mesma transformação linear. Matrizes semelhantes 123
- 4.10. Exercícios 127

5. VALORES PRÓPRIOS DE OPERADORES EM ESPAÇOS EUCLIDIANOS

- 5.1. Valores próprios e productos internos 129
- 5.2. Transformações hermiticas e hemi-hermiticas 130
- 5.3. Valores próprios e vectores próprios de operadores hermiticos e hemi-hermiticos 132
- 5.4. Ortogonalidade de vectores próprios correspondentes a valores próprios distintos 133
- 5.5. Exercícios 134
- 5.6. Existência de um conjunto ortonormal de vectores próprios para operadores hermiticos e hemi-hermiticos em espaços de dimensão finita 135
- 5.7. Representação matricial de operadores hermiticos e hemi-hermiticos 137
- 5.8. Matrizes hermiticas e hemi-hermiticas. A associada de uma matriz 138
- 5.9. Diagonalização de uma matriz hermitica ou hemi-hermitica 138
- 5.10. Matrizes unitárias. Matrizes ortogonais 139
- 5.11. Exercícios 140
- 5.12. Formas quadráticas 143
- 5.13. Redução de uma forma quadrática real à forma diagonal 145
- 5.14. Aplicações à geometria analítica 147
- 5.15. Exercícios 151
- *5.16. Valores próprios de uma transformação simétrica obtidos como valores de sua forma quadrática 152
- *5.17. Propriedades extremas dos valores próprios de uma transformação simétrica 154
- *5.18. O caso de dimensão finita 155
- 5.19. Transformações unitárias 155
- 5.20. Exercícios 158

6. EQUAÇÕES DIFERENCIAIS LINEARES

- 6.1. Introdução histórica 161
- 6.2. Revisão dos resultados já estabelecidos relativos às equações diferenciais lineares de primeira e de segunda ordem 162
- 6.3. Exercícios 164
- 6.4. Equações diferenciais lineares de ordem n 165
- 6.5. O teorema de existência e unicidade 166
- 6.6. A dimensão do espaço solução de uma equação linear homogénea 167

6.7.	A álgebra de operadores de coeficientes constantes	168
6.8.	Determinação de uma base de soluções para equações lineares com coeficientes constantes por factorização de operadores	170
6.9.	Exercícios	175
6.10.	Relação entre as equações homogêneas e não homogêneas	177
6.11.	Determinação de uma solução particular da equação não homogênea. O método de variação das constantes	178
6.12.	Não singularidade da matriz wronskiana de n soluções independentes de uma equação linear homogênea	182
6.13.	Métodos especiais para determinação de soluções particulares de equações não homogêneas. Redução a um sistema de equações lineares de primeira ordem	184
6.14.	O método do anulador para determinação de uma solução particular da equação não homogênea	185
6.15.	Exercícios	188
6.16.	Exercícios variados sobre equações diferenciais lineares	189
6.17.	Equações lineares de segunda ordem com coeficientes analíticos	191
6.18.	A equação de Legendre	194
6.19.	Os polinómios de Legendre	197
6.20.	Fórmula de Rodrigues para os polinómios de Legendre	199
6.21.	Exercícios	200
6.22.	O método de Frobenius	204
6.23.	A equação de Bessel	206
6.24.	Exercícios	212

7. SISTEMAS DE EQUAÇÕES DIFERENCIAIS

7.1.	Introdução	215
7.2.	Conceitos do cálculo para funções matriciais	218
7.3.	Séries de matrizes. Normas de matrizes	218
7.4.	Exercícios	220
7.5.	A matriz exponencial	221
7.6.	A equação diferencial verificada por e^{tA}	222
7.7.	Teorema da unicidade para a equação diferencial matricial $F'(t) = AF(t)$	223
7.8.	Regra do producto de exponenciais de matrizes	224
7.9.	Teoremas de existência e unicidade para sistemas lineares homogêneos	225
7.10.	O problema do cálculo de e^{tA}	226
7.11.	O teorema de Cayley-Hamilton	228
7.12.	Exercícios	230
7.13.	Método de Putzer para o cálculo de e^{tA}	231
7.14.	Outros métodos para calcular e^{tA} em casos particulares	235
7.15.	Exercícios	238
7.16.	Sistemas lineares não homogêneas com coeficientes constantes	239
7.17.	Exercícios	241
7.18.	O sistema linear geral $Y'(t) = P(t)Y(t) + Q(t)$	244
7.19.	Resolução de sistemas lineares homogêneos por intermédio de séries de potências	248
7.20.	Exercícios	249

7.21.	Demonstração do teorema de existência pelo método das aproximações sucessivas	258
7.22.	O método das aproximações sucessivas aplicado a sistemas não lineares de primeira ordem	255
7.23.	Demonstração de um teorema de existência e unicidade para sistemas não lineares de primeira ordem	257
7.24.	Exercícios	259
*7.25.	Aproximações sucessivas e pontos fixos de operadores	261
*7.26.	Espaços lineares normados	262
*7.27.	Operadores de contracção	263
*7.28.	Teorema do ponto fixo para operadores de contracção	264
*7.29.	Aplicações do teorema do ponto fixo	266

PARTE 2. ANÁLISE NÃO LINEAR

8. CÁLCULO DIFERENCIAL EM CAMPOS ESCALARES E VECTORIAIS

8.1.	Funções de R^n em R^m . Campos vectoriais e escalares	273
8.2.	Bolas abertas e conjuntos abertos	274
8.3.	Exercícios	276
8.4.	Limites e continuidade	278
8.5.	Exercícios	282
8.6.	A derivada de um campo escalar relativamente a um vector	283
8.7.	Derivadas direccionais e derivadas parciais	286
8.8.	Derivadas parciais de ordem superior	287
8.9.	Exercícios	287
8.10.	Derivadas direccionais e continuidade	288
8.11.	A diferencial	290
8.12.	Gradiente de um campo escalar	291
8.13.	Uma condição suficiente de diferenciabilidade	293
8.14.	Exercícios	295
8.15.	Generalização da regra de derivação de funções compostas para derivadas de campos escalares	296
8.16.	Aplicações geométricas. Conjuntos de nível. Planos tangentes	298
8.17.	Exercícios	301
8.18.	Derivadas de campos vectoriais	303
8.19.	A diferenciabilidade implica a continuidade	304
8.20.	Generalização da regra de derivação da função composta para derivadas de campos vectoriais	305
8.21.	Forma matricial da regra de derivação para a composição	306
8.22.	Exercícios	309
*8.23.	Condições suficientes para a igualdade das derivadas parciais mistas	311
8.24.	Exercícios variados	315

9. APLICAÇÕES DO CÁLCULO DIFERENCIAL

9.1.	Equações de derivadas parciais	319
------	--------------------------------	-----

9.2.	Uma equação de derivadas parciais de primeira ordem com coeficientes constantes	320	
9.3.	Exercícios	322	
9.4.	A equação unidimensional das ondas	324	
9.5.	Exercícios	329	
9.6.	Derivadas de funções implícitas	331	
9.7.	Exemplos resolvidos	335	
9.8.	Exercícios	340	
9.9.	Máximos, mínimos e pontos sela	341	
9.10.	Fórmula de Taylor de segunda ordem para campos escalares	346	
9.11.	A natureza do ponto de estacionaridade determinada pelos valores próprios da matriz Hessiana	348	
9.12.	Crítério das derivadas de segunda ordem para extremos de funções de duas variáveis	351	
9.13.	Exercícios	351	
9.14.	Extremos condicionados. Multiplicadores de Lagrange	353	
9.15.	Exercícios	357	
9.16.	Teorema do valor extremo para campos escalares contínuos	358	
9.17.	O teorema da continuidade uniforme para campos escalares contínuos	361	

10. INTEGRAIS DE LINHA

10.1.	Introdução	363	
10.2.	Integrais de linha e linhas de integração	363	
10.3.	Outras notações para os integrais de linha	364	
10.4.	Propriedades fundamentais dos integrais de linha	366	
10.5.	Exercícios	368	
10.6.	O conceito de trabalho como um integral de linha	369	
10.7.	Integrais de linha relativos ao comprimento de arco	370	
10.8.	Outras aplicações dos integrais de linha	371	
10.9.	Exercícios	372	
10.10.	Conjuntos conexos abertos. Independência da linha	374	
10.11.	O segundo teorema fundamental do cálculo para as integrais de linha	374	
10.12.	Aplicações à mecânica	376	
10.13.	Exercícios	378	
10.14.	O primeiro teorema fundamental do cálculo para integrais de linha	379	
10.15.	Condições necessárias e suficientes para que um campo de vectores seja um gradiente	381	
10.16.	Condições necessárias para que um campo vectorial seja um gradiente	382	
10.17.	Métodos especiais de construção de funções potenciais	384	
10.18.	Exercícios	387	
10.19.	Aplicações às equações diferenciais exactas de primeira ordem	389	
10.20.	Exercícios	392	
10.21.	Funções potenciais em conjuntos convexos	393	

11. INTEGRAIS MÚLTIPLOS

11.1.	Introdução	397
-------	------------	-----

11.2.	Partições de retângulos. Funções em escada	398	
11.3.	O integral duplo de uma função em escada	399	
11.4.	A definição de integral duplo de uma função definida e limitada num retângulo	401	
11.5.	Integrais duplos superior e inferior	402	
11.6.	Cálculo de um integral duplo por integração unidimensional repetida		403
11.7.	Interpretação geométrica do integral duplo como um volume	404	
11.8.	Exemplos resolvidos	405	
11.9.	Exercícios	407	
11.10.	Integrabilidade de funções contínuas	408	
11.11.	Integrabilidade de funções limitadas com descontinuidades		409
11.12.	Integrais duplos estendidos a regiões mais gerais	410	
11.13.	Aplicações a áreas e volumes	414	
11.14.	Exemplos resolvidos	415	
11.15.	Exercícios	417	
11.16.	Outras aplicações dos integrais duplos	419	
11.17.	Dois teoremas de Pappus	422	
11.18.	Exercícios	424	
11.19.	Teorema de Green no plano	425	
11.20.	Algumas aplicações do teorema de Green	429	
11.21.	Uma condição necessária e suficiente para que um campo vectorial bidimensional seja um gradiente	430	
11.22.	Exercícios	433	
*11.23.	Teorema de Green para regiões multiplamente conexas	435	
*11.24.	O número de giros	437	
*11.25.	Exercícios	439	
11.26.	Mudança de variáveis num integral duplo	441	
11.27.	Casos particulares da fórmula de mudança de variáveis	445	
11.28.	Exercícios	449	
11.29.	Demonstração da fórmula de mudança de variáveis num caso particular	450	
11.30.	Demonstração da fórmula de mudança de variáveis no caso geral		453
11.31.	Extensões a um número superior de dimensões	455	
11.32.	Mudança de variáveis num integral n -múltiplo	457	
11.33.	Exemplos resolvidos	459	
11.34.	Exercícios	463	

12. INTEGRAIS DE SUPERFÍCIE

12.1.	Representação paramétrica de uma superfície	467
12.2.	O producto vectorial fundamental	471
12.3.	O producto vectorial fundamental definido uma normal à superfície	474
12.4.	Exercícios	475
12.5.	Área de uma superfície na representação paramétrica	475
12.6.	Exercícios	481
12.7.	Integrais de superfície	481
12.8.	Mudança de representação paramétrica	484
12.9.	Outras notações para os integrais de superfície	486

12.10.	Exercícios	489	
12.11.	O teorema de Stokes	490	
12.12.	O rotacional e a divergência de um campo vectorial	493	
12.13.	Exercícios	495	
12.14.	Outras propriedades do rotacional e da divergência	496	
12.15.	Exercícios	500	
*12.16.	Reconstrução de um campo vectorial a partir do seu rotacional	502	
*12.17.	Exercícios	506	
12.18.	Extensões do teorema de Stokes	507	
12.19.	O teorema da divergência (teorema de Gauss)	511	
12.20.	Aplicações do teorema da divergência	515	
12.21.	Exercícios	517	

PARTE 3. TEMAS ESPECIAIS

13. FUNÇÕES DE CONJUNTO E PROBABILIDADE ELEMENTAR

13.1.	Introdução histórica	525	
13.2.	Funções de conjunto completamente aditivas	526	
13.3.	Medidas finitamente aditivas	528	
13.4.	Exercícios	529	
13.5.	A definição de probabilidade para conjuntos fundamentais finitos	530	
13.6.	Terminologia peculiar da teoria das probabilidades	533	
13.7.	Exercícios	534	
13.8.	Exemplos resolvidos	535	
13.9.	Exercícios	537	
13.10.	Alguns princípios básicos de análise combinatória	539	
13.11.	Exercícios	544	
13.12.	Probabilidade condicionada	545	
13.13.	Independência aleatória	547	
13.14.	Exercícios	549	
13.15.	Experiências compostas	551	
13.16.	Esquema de Bernoulli	555	
13.17.	O número mais favorável de ocorrências do acontecimento favorável em n experiências dum esquema de Bernoulli	557	
13.18.	Exercícios	560	
13.19.	Conjuntos numeráveis e não numeráveis	562	
13.20.	Exercícios	566	
13.21.	Definição de probabilidade para conjuntos fundamentais infinitos numeráveis	567	
13.22.	Exercícios	569	
13.23.	Exercícios variados sobre probabilidades	569	

14. CALCULO DAS PROBABILIDADES

14.1.	A definição de probabilidade para conjuntos fundamentais não numeráveis	573	
14.2.	Numerabilidade de conjuntos de pontos com probabilidade positiva	574	

14.3.	Variáveis aleatórias	575
14.4.	Exercícios	577
14.5.	Funções de repartição	578
14.6.	Discontinuidades das funções de repartição	582
14.7.	Distribuições discretas. Funções de massa probabilística	585
14.8.	Exercícios	588
14.9.	Distribuições contínuas. Funções densidade	591
14.10.	Distribuição uniforme num intervalo	592
14.11.	Distribuição de Cauchy	597
14.12.	Exercícios	598
14.13.	Distribuições exponenciais	599
14.14.	Distribuições normais	602
14.15.	Indicações referentes a distribuições mais gerais	606
14.16.	Exercícios	607
14.17.	Distribuições de funções de variáveis aleatórias	608
14.18.	Exercícios	609
14.19.	Distribuição de variáveis aleatórias bidimensionais	610
14.20.	Distribuições discretas bidimensionais	613
14.21.	Distribuições bidimensionais contínuas. Funções densidade	614
14.22.	Exercícios	616
14.23.	Distribuição de funções de duas variáveis aleatórias	618
14.24.	Exercícios	622
14.25.	Esperança matemática e variância	625
14.26.	Esperança matemática de uma função de uma variável aleatória	629
14.27.	Exercícios	630
14.28.	Desigualdade de Tchebycheff	632
14.29.	Leis dos grandes números	634
14.30.	Teorema limite central do cálculo das probabilidades	637
14.31.	Exercícios	639
	Bibliografia	641

15. INTRODUÇÃO A ANÁLISE NUMÉRICA

15.1.	Introdução histórica	643
15.2.	Aproximação polinomial	644
15.3.	Aproximação polinomial e espaços lineares normados	646
15.4.	Problemas fundamentais da aproximação polinomial	648
15.5.	Exercícios	650
15.6.	Polinómios interpoladores	652
15.7.	Pontos de interpolação igualmente separados	655
15.8.	Análise do erro na interpolação polinomial	656
15.9.	Exercícios	659
15.10.	Fórmula de interpolação de Newton	662
15.11.	Pontos de interpolação igualmente espaçados. O operador das diferenças sucessivas	664
15.12.	Polinómios factoriais	666
15.13.	Exercícios	667
15.14.	Um problema de número relativo à norma maximal	669

15.15.	Polinómios de Tchebycheff	670	
15.16.	Uma propriedade de mínimo dos polinómios de Tchebycheff	672	
15.17.	Aplicação á formula de erro na interpolação	674	
15.18.	Exercícios	675	
15.19.	Integração aproximada. A regra trapezoidal	677	
15.20.	Regra de Simpson	680	
15.21.	Exercícios	685	
15.22.	A fórmula de somação de Euler	688	
15.23.	Exercícios	694	
	Bibliografia	697	
	Soluções dos exercícios	699	
	Índice alfabético	745	

Cálculo

PARTE 1
ANÁLISE LINEAR

1

ESPAÇOS LINEARES

1.1. Introdução

No desenvolvimento da Matemática encontramos muitos exemplos de objectos matemáticos que podem ser adicionados uns aos outros e multiplicados por números reais. O primeiro exemplo de tais objectos são os próprios números reais. Outros exemplos são as funções reais, os números complexos, as séries infinitas, os vectores num espaço n dimensional e as funções vectoriais. Neste capítulo vamos analisar um conceito matemático geral, chamado *espaço linear*, que inclui todos estes exemplos e muitos outros como casos particulares.

Em resumo, um espaço linear é um conjunto de elementos de natureza qualquer no qual se efectuam certas operações (chamadas *adição* e *multiplicação por números*). Ao definir-se um espaço linear, *não é necessário especificar* a natureza dos elementos nem dizer como se realizam entre eles as operações acabadas de referir. Em vez disso, exige-se que as operações gozem de certas propriedades que se tomam como axiomas do espaço linear. Vamos precisamente, em seguida, fazer uma descrição pormenorizada desses axiomas.

1.2. Definição de espaço linear

Seja V um conjunto não vazio de objectos, chamados *elementos*. O conjunto V chama-se um espaço linear se satisfaz aos dez axiomas que a seguir se enunciam, divididos em três grupos.

Axiomas de fecho.

AXIOMA 1. FECHO A RESPECTO DA ADIÇÃO. *A todo o par de elementos x e y de V corresponde um único elemento de V , chamado soma de x e y e representado por $x + y$.*

AXIOMA 2. FECHO A RESPEITO DA MULTIPLICAÇÃO POR NÚMEROS REAIS. *A todo o x de V e todo o número real a corresponde um elemento de V , chamado o produto de a por x e representado por ax .*

Axiomas para a adição.

AXIOMA 3. PROPRIEDADE COMUTATIVA. *Para todo o x e y de V , tem-se $x + y = y + x$.*

AXIOMA 4. PROPRIEDADE ASSOCIATIVA. *Para todo o x , y e z de V , tem-se $x + (y + z) = (x + y) + z$.*

AXIOMA 5. EXISTÊNCIA DE ELEMENTO ZERO. *Existe um elemento em V , representado pelo símbolo O , tal que*

$$x + O = x \quad \text{para todo o } x \text{ de } V.$$

AXIOMA 6. EXISTÊNCIA DE SIMÉTRICOS. *Para todo o x de V , o elemento $(-1)x$ tem a propriedade*

$$x + (-1)x = O.$$

Axiomas para a multiplicação por números.

AXIOMA 7. PROPRIEDADE ASSOCIATIVA. *Para todo o x de V , e todo o par de números reais a e b , tem-se*

$$a(bx) = (ab)x.$$

AXIOMA 8. PROPRIEDADE DISTRIBUTIVA PARA A ADIÇÃO EM V . *Para todo o par x e y de V e todo o real a , tem-se*

$$a(x + y) = ax + ay.$$

AXIOMA 9. PROPRIEDADE DISTRIBUTIVA PARA A ADIÇÃO DE NÚMEROS. *Para todo o x em V e todo o par de reais a e b tem-se*

$$(a + b)x = ax + bx.$$

AXIOMA 10. EXISTÊNCIA DE ELEMENTO IDENTIDADE. *Para todo x em V , tem-se $1x = x$.*

Os espaços lineares, como foram definidos atrás, são muitas vezes chamados espaços lineares *reais*, para fazer ressaltar o facto de que se multiplicam elementos de V por números reais. Se nos Axiomas 2, 7, 8 e 9 substituímos *número real* por *número*

complexo, a estrutura resultante chama-se um *espaço linear complexo*. Por vezes um espaço linear chama-se também *espaço vectorial linear*, ou mais simplesmente *espaço vectorial*; os números usados como multiplicadores chamam-se *escalares*. Um espaço linear real admite os números reais como escalares, um espaço linear complexo admite os números complexos como escalares. Embora se considerem aqui fundamentalmente exemplos de espaços vectoriais lineares reais, todos os teoremas são verdadeiros igualmente para os espaços vectoriais complexos. Quando fazemos uso da expressão espaço linear, sem qualquer designação suplementar, deve subentender-se que o espaço pode ser real ou complexo.

1.3. Exemplos de espaços lineares

Se especificamos qual o conjunto V e dizemos como somar os seus elementos e como multiplicá-los por números, obtemos um exemplo concreto de um espaço linear. O leitor pode facilmente verificar que cada um dos seguintes exemplos satisfaz a todos os axiomas para um espaço linear real.

EXEMPLO 1. Seja $V = \mathbf{R}$ o conjunto dos números reais, e sejam $x + y$ e ax a adição e multiplicação usuais de números reais.

EXEMPLO 2. Seja $V = \mathbf{C}$ o conjunto dos números complexos, e seja $x + y$ a adição ordinária de números complexos e ax a multiplicação de números complexos x pelo número real a . Embora os elementos de V sejam números complexos, este é um espaço linear real porque os escalares são reais.

EXEMPLO 3. Seja $V = V_n$ o espaço vectorial dos sistemas de n números reais, com a adição e a multiplicação por escalares definida da maneira usual em função das componentes.

EXEMPLO 4. Seja V o conjunto de todos os vectores em V_n ortogonais a um dado vector não nulo N . Se $n = 2$, este espaço linear é uma recta que passa por O , admitindo N como vector normal. Se $n = 3$, é um plano que passa por O com N como vector normal.

Os exemplos que se seguem dizem-se *espaços funcionais*. Os elementos de V são funções reais, com a adição de duas funções f e g definidas na forma usual:

$$(f + g)(x) = f(x) + g(x)$$

para todo o real x pertencente à intersecção dos domínios de f e g . A multiplicação de uma função f por um escalar real a define-se do modo seguinte: af é a função cujo valor para cada x no domínio de f é $af(x)$. O elemento zero é a função cujos valores são sempre zero. O leitor verificará com facilidade que cada um dos conjuntos seguintes é um espaço funcional.

EXEMPLO 5. O conjunto de todas as funções definidas num dado intervalo.

EXEMPLO 6. O conjunto de todos os polinómios.

EXEMPLO 7. O conjunto de todos os polinómios de grau $\leq n$, com n fixo. (Sempre que se considera este conjunto subentende-se que o polinómio zero está também incluído). O conjunto de todos os polinómios de grau *igual a* n não é um espaço linear porque os axiomas de fecho não são satisfeitos. Por exemplo, a soma de dois polinómios de grau n não terá necessariamente grau n .

EXEMPLO 8. O conjunto de todas as funções contínuas num dado intervalo. Se o intervalo é $[a, b]$ representamos este espaço linear por $C(a, b)$.

EXEMPLO 9. O conjunto de todas as funções deriváveis num dado ponto.

EXEMPLO 10. O conjunto de todas as funções integráveis num dado intervalo.

EXEMPLO 11. O conjunto de todas as funções f definidas no ponto 1, com $f(1) = 0$. O número 0 é fundamental neste exemplo. Se substituirmos 0 por um número c não nulo, violamos os axiomas de fecho.

EXEMPLO 12. O conjunto de todas as soluções de uma equação diferencial linear homogénea $y'' + ay' + by = 0$, com a e b constantes. Aqui mais uma vez o 0 é essencial. O conjunto de soluções de uma equação diferencial não homogénea não satisfaz aos axiomas de fecho.

Estes exemplos e muitos outros mostram bem quanto o conceito de espaço linear está estendido à Álgebra, Geometria e Análise. Quando se deduz um teorema a partir dos axiomas de um espaço linear, obtemos, de uma vez, um resultado válido para cada exemplo concreto. Unificando diferentes exemplos desta maneira ganhamos um conhecimento mais aprofundado de cada um. Algumas vezes o conhecimento de um exemplo particular ajuda-nos a antecipar ou interpretar resultados válidos para outros exemplos e põe em evidência relações que de outro modo poderiam passar despercebidas.

1.4. Consequências elementares dos axiomas

Os teoremas que se seguem deduzem-se facilmente dos axiomas para um espaço linear.

TEOREMA 1.1. UNICIDADE DO ELEMENTO ZERO. *Em qualquer espaço linear existe um e um só elemento zero.*

Demonstração. O axioma 5 diz-nos que existe pelo menos um elemento zero. Suponhamos que existiam dois, por exemplo O_1 e O_2 . Tomando $x = O_1$ e $O = O_2$ no Axioma

5, obtemos $O_1 + O_2 = O_1$. Analogamente, tomando $x = O_2$ e $O = O_1$, encontramos $O_2 + O_1 = O_2$. Mas $O_1 + O_2 = O_2 + O_1$, devido à propriedade comutativa, pelo que $O_1 = O_2$.

TEOREMA 1.2. UNICIDADE DOS ELEMENTOS SIMÉTRICOS. *Em qualquer espaço linear todo o elemento admite unicamente um simétrico, isto é, para todo o x existe um e um só y tal que $x + y = O$.*

Demonstração. O Axioma 6 diz-nos que cada x admite pelo menos um simétrico, a saber $(-1)x$. Admitamos agora que x tinha dois simétricos, y_1 e y_2 . Então $x + y_1 = O$ e $x + y_2 = O$. Somando y_2 a ambos os membros da primeira igualdade e utilizando os Axiomas 5, 4 e 3, encontramos

$$y_2 + (x + y_1) = y_2 + O = y_2,$$

e

$$y_2 + (x + y_1) = (y_2 + x) + y_1 = O + y_1 = y_1 + O = y_1.$$

Portanto $y_1 = y_2$, pelo que x tem precisamente um simétrico, o elemento $(-1)x$.

Notação. O simétrico de x representa-se por $-x$. A diferença $y - x$ é definida pela soma $y + (-x)$.

O teorema seguinte refere um certo número de propriedades que regem os cálculos algébricos elementares num espaço linear.

TEOREMA 1.3. *Num dado espaço linear, sejam x e y elementos arbitrários e a e b escalares arbitrários. Então verificam-se as seguintes propriedades:*

(a) $0x = O$.

(b) $aO = O$.

(c) $(-a)x = -(ax) = a(-x)$.

(d) Se $ax = O$, então ou $a = 0$ ou $x = O$.

(e) Se $ax = ay$ e $a \neq 0$, então $x = y$.

(f) Se $ax = bx$ e $x \neq O$, então $a = b$.

(g) $-(x + y) = (-x) + (-y) = -x - y$.

(h) $x + x = 2x$, $x + x + x = 3x$, e em geral $\sum_{i=1}^n x = nx$.

Vamos demonstrar (a), (b) e (c), deixando as demonstrações das restantes ao cuidado do leitor.

Demonstração de (a). Seja $z = 0x$. Desejamos provar que $z = O$. Somando z a si próprio e aplicando o Axioma 9, verificamos que

$$z + z = 0x + 0x = (0 + 0)x = 0x = z.$$

Adicionamos agora $-z$ a ambos os membros e obtemos $z = O$.

Demonstração de (b). Seja $z = aO$, adicionemos z a si próprio e utilizemos o Axioma 8.

Demonstração de (c). $z = (-a)x$. Adicionando z a ax e utilizando o Axioma 9, verificamos que

$$z + ax = (-a)x + ax = (-a + a)x = 0x = O,$$

pelo que z é o simétrico de ax , $z = -(ax)$. Analogamente, se adicionamos $a(-x)$ a ax e utilizamos o Axioma 8 e a propriedade (b), encontramos que $a(-x) = -(ax)$.

1.5. Exercícios

Nos Exercícios 1 a 28, determinar se cada um dos conjuntos dados é um espaço linear real, com a adição e a multiplicação por escalares reais definidas da forma usual. Para os Exercícios em que assim não seja, dizer quais são os axiomas que não se verificam. As funções nos Exercícios 1 a 17 são reais. Nos Exercícios 3, 4 e 5 cada função tem um domínio contendo 0 e 1. Nos Exercícios 7 a 12, o domínio é o conjunto de todos os números reais.

1. Todas as funções racionais.
2. Todas as funções racionais f/g , com o grau de $f \leq$ que o grau de g (incluindo $f = 0$).
3. Todas as funções f com $f(0) = f(1)$.
4. Todas as funções f com $2f(0) = f'(1)$.
5. Todas as funções f com $f(1) = 1 + f(0)$.
6. Todas as funções em escada definidas em escada $[0, 1]$.
7. Todas as funções com $f(x) \rightarrow 0$ quando $x \rightarrow +\infty$.
8. Todas as funções pares.
9. Todas as funções ímpares.
10. Todas as funções limitadas.
11. Todas as funções crescentes.
12. Todas as funções periódicas de período 2π .
13. Todas as funções f integráveis em $[0, 1]$, com $\int_0^1 f(x)dx = 0$.
14. Todas as funções f integráveis em $[0, 1]$, com $\int_0^1 f(x)dx \geq 0$.
15. Todas as funções verificando $f(x) = f(1 - x)$ para todo x .
16. Todos os polinômios de Taylor de grau $\leq n$ para um n dado (incluindo o polinômio zero).
17. Todas as soluções da equação diferencial linear homogênea de segunda ordem $y'' + P(x)y' + Q(x)y = 0$, com P e Q funções dadas e contínuas para todo x .
18. Todas as sucessões reais limitadas.
19. Todas as sucessões reais convergentes.
20. Todas as séries reais convergentes.
21. Todas as séries reais absolutamente convergentes.
22. Todos os vectores (x, y, z) de V_3 com $z = 0$.
23. Todos os vectores (x, y, z) de V_3 com $x = 0$ ou $y = 0$.
24. Todos os vectores (x, y, z) de V_3 com $y = 5x$.
25. Todos os vectores (x, y, z) de V_3 com $3x + 4y = 1, z = 0$.
26. Todos os vectores (x, y, z) de V_3 que são múltiplos escalares de $(1, 2, 3)$.
27. Todos os vectores (x, y, z) de V_3 cujas componentes satisfazem a um sistema de três equações lineares de forma:

$$a_{11}x + a_{12}y + a_{13}z = 0, \quad a_{21}x + a_{22}y + a_{23}z = 0, \quad a_{31}x + a_{32}y + a_{33}z = 0.$$

28. Todos os vectores de V_n que são combinações lineares de dois vectores dados A e B .
29. Seja $V = \mathbb{R}^+$, o conjunto dos números reais positivos. Defina-se a "soma" de dois elementos x e y em V como sendo o seu produto $x \cdot y$ (no sentido usual) e defina-se "multiplicação" de um elemento x de V por um escalar c como sendo cx . Provar que V é um espaço linear real com 1 como elemento zero.
30. (a) Provar que o Axioma 10 pode ser provado a partir dos outros axiomas.
 (b) Provar que o Axioma 10 não pode ser deduzido dos outros Axiomas se o Axioma 6 for substituído pelo Axioma 6': "Para todo x em V existe um elemento y de V tal que $x + y = O$ ".
31. Seja S o conjunto de todos os pares ordenados (x_1, x_2) de números reais. Em cada alínea determinar se sim ou não S é um espaço linear com as operações de adição e multiplicação por escalares definidas como se indica. Se o conjunto não for um espaço linear, dizer quais os axiomas que não são verificados.
- (a) $(x_1, x_2) + (y_1, y_2) = (x_1 + y_1, x_2 + y_2)$, $a(x_1, x_2) = (ax_1, 0)$.
 (b) $(x_1, x_2) + (y_1, y_2) = (x_1 + y_1, 0)$, $a(x_1, x_2) = (ax_1, ax_2)$.
 (c) $(x_1, x_2) + (y_1, y_2) = (x_1, x_2 + y_2)$, $a(x_1, x_2) = (ax_1, ax_2)$.
 (d) $(x_1, x_2) + (y_1, y_2) = (|x_1 + x_2|, |y_1 + y_2|)$, $a(x_1, x_2) = (|ax_1|, |ax_2|)$.
32. Demonstrar as partes da d até h do Teorema 1.3.

1.6. Subespaços de um espaço linear

Dado um espaço linear V , seja S um conjunto não vazio de V . Se S é também um espaço linear, com as mesmas operações de adição e multiplicação por escalares, então S diz-se um *subespaço* de V . O teorema que apresentamos a seguir dá um critério simples para determinar se sim ou não um subconjunto de um espaço linear é um subespaço.

TEOREMA 1.4. *Se S é um subconjunto não vazio de um espaço linear V , então S é um subespaço se e só se S satisfaz aos axiomas de fecho.*

Demonstração. Se S é um subespaço, verificam-se todos os axiomas para um espaço linear e por conseguinte, em particular, verificam-se os axiomas de fecho.

Demonstremos agora que se S satisfaz aos axiomas de fecho, satisfaz igualmente aos outros. As propriedades comutativa e associativa para a adição (Axiomas 3 e 4) e os axiomas para a multiplicação por escalares (Axiomas 7 e 10) são automaticamente satisfeitos em S porque são válidos para todos os elementos de V . Faltam verificar os Axiomas 5 e 6, a existência em S do elemento zero e a existência do simétrico de cada elemento de S .

Seja x um qualquer elemento de S . (S tem pelo menos um elemento visto que é não vazio.) Pelo Axioma 2, ax está em S para todo o escalar a . Fazendo $a = 0$, resulta que $0x$ está em S . Mas $0x = O$, pelo teorema 1.3(a), pelo que $O \in S$ e o Axioma 5 é satisfeito. Fazendo $a = -1$, vemos que $(-1)x$ pertence a S . Mas $x + (-1)x = O$ visto que quer x , quer $(-1)x$ estão em V , e consequentemente o Axioma 6 é satisfeito em S . Deste modo S é subespaço de V .

DEFINIÇÃO. *Seja S um subconjunto não vazio de um espaço linear V . Um elemento x de V da forma*

$$x = \sum_{i=1}^k c_i x_i,$$

onde x_1, x_2, \dots, x_k pertencem todos a S e c_1, c_2, \dots, c_k são escalares, diz-se uma combinação linear finita de elementos de S . O conjunto de todas as combinações lineares finitas de elementos de S verificam os axiomas de fecho e por conseguinte é um subespaço de V . Chama-se este o subespaço gerado por S , e representa-se por $L(S)$. Se S é vazio, definimos $L(S)$ como $\{O\}$, o conjunto constando unicamente do elemento zero.

Conjuntos distintos podem gerar o mesmo subespaço. Por exemplo, o espaço V_2 é gerado por cada um dos seguintes conjuntos de vectores $\{i, j\}$, $\{i, j, i+j\}$, $\{O, i-i, j-j, i+j\}$. O espaço de todos os polinómios $np(t)$ de grau $\leq n$ é gerado pelo conjunto de $n+1$ polinómios.

$$\{1, t, t^2, \dots, t^n\}.$$

É também gerado pelo conjunto $\{1, t/2, t^2/3, \dots, t^n/(n+1)\}$, e por $\{1, (1+t), (1+t)^2, \dots, (1+t)^n\}$. O espaço de todos os polinómios é gerado pelo conjunto infinito dos polinómios $\{1, t, t^2, \dots\}$.

Um certo número de perguntas se podem pôr ao chegarmos a este ponto. Por exemplo, que espaços podem ser gerados por um conjunto finito de elementos? Se um espaço pôde ser gerado por um conjunto finito de elementos, qual é o número mínimo de elementos necessários? Para analisar estas e outras questões, introduzimos os conceitos de *dependência* e *independência linear*, *bases* e *dimensão*. Estas noções já foram referidas no capítulo 12 quando do estudo do espaço vectorial V_n . Agora apenas as vamos generalizar aos espaços lineares de tipo qualquer.

1.7. Conjuntos dependentes e independentes num espaço linear

DEFINIÇÃO. Um conjunto S de elementos de um espaço linear V diz-se *dependente* se existe um conjunto finito de elementos distintos de S ; por exemplo x_1, x_2, \dots, x_k e um correspondente conjunto de escalares c_1, c_2, \dots, c_k , não conjuntamente todos nulos, tais que

$$\sum_{i=1}^k c_i x_i = O.$$

Uma equação $\sum_{i=1}^k c_i x_i = O$ com algum $c_i \neq 0$ diz-se ser uma representação não trivial de O . O conjunto S diz-se *linearmente independente* se não é dependente, isto é, quaisquer que sejam os elementos distintos x_1, x_2, \dots, x_k de S e os escalares c_1, c_2, \dots, c_k ,

$$\sum_{i=1}^k c_i x_i = O \quad \text{implica} \quad c_1 = c_2 = \dots = c_k = 0.$$

Embora a dependência e independência sejam propriedades dos conjuntos de elementos, aplicam-se habitualmente estas designações aos próprios elementos desses

mesmos conjuntos. Por exemplo, os elementos de um conjunto independente dizem-se linearmente independentes.

Se S é um conjunto finito, a definição precedente concorda com a dada no Capítulo 12 para o espaço V_n . Contudo, a presente definição não está restringida a conjuntos finitos.

EXEMPLO 1. Se um subconjunto T de um conjunto S é dependente, então S também é *dependente*. Isto é logicamente equivalente à afirmação de que cada subconjunto de um conjunto independente é independente.

EXEMPLO 2. Se um elemento de S é um múltiplo escalar do outro, então S é dependente.

EXEMPLO 3. Se $0 \in S$, então S é dependente.

EXEMPLO 4. O conjunto vazio é independente.

No Volume I foram discutidos muitos exemplos de conjuntos dependentes e independentes de vectores de V_n . Os exemplos seguintes ilustram esses conceitos em espaços funcionais. Em cada caso o espaço linear fundamentalmente V é o conjunto de todas as funções reais definidas na recta real.

EXEMPLO 5. Sejam $u_1(t) = \cos^2 t$, $u_2(t) = \sin^2 t$, $u_3(t) = 1$, para todo o número real t . A identidade de Pitágoras mostra que $u_1 + u_2 - u_3 = 0$, pelo que as três funções u_1, u_2, u_3 , são dependentes.

EXEMPLO 6. Seja $u_k(t) = t^k$ para $k = 0, 1, 2, \dots$, e t real. O conjunto $S = \{u_0, u_1, \dots\}$ é independente. Para demonstrar isto, basta provar que para cada n , os $n + 1$ polinómios u_0, u_1, \dots, u_n são independentes. Uma relação da forma $\sum c_k u_k = 0$ significa que

$$(1.1) \quad \sum_{k=0}^n c_k t^k = 0$$

para todo o real t . Quando $t = 0$, encontramos $c_0 = 0$. Derivando (1.1) e fazendo $t = 0$, encontramos $c_1 = 0$. Repetindo o processo, verificamos que cada coeficiente c_k é zero.

EXEMPLO 7. Se a_1, \dots, a_n são números reais distintos, as n funções exponenciais

$$u_1(x) = e^{a_1 x}, \dots, u_n(x) = e^{a_n x}$$

são independentes. Podemos demonstrá-lo por indução relativamente a n . O resultado verifica-se trivialmente quando $n = 1$. Admitamos por conseguinte que é verdadeira para $n - 1$ funções exponenciais e consideremos os escalares c_1, c_1, \dots, c_n tais que

$$(1.2) \quad \sum_{k=1}^n c_k e^{a_k x} = 0.$$

Seja a_M o maior dos n números a_1, a_2, \dots, a_n . Multiplicando ambos os membros de (1.2) por $e^{-a_M x}$, obtemos

$$(1.3) \quad \sum_{k=1}^n c_k e^{(a_k - a_M)x} = 0.$$

Se $k \neq M$, o número $a_k - a_M$ é negativo. Deste modo, quando $x \rightarrow +\infty$ na equação (1.3), cada termo com $k \neq M$ tende para zero e encontramos que $c_M = 0$. Suprimindo o termo de ordem M em (1.2) e aplicando a hipótese de indução, encontramos que cada um dos $n - 1$ restantes coeficientes c_k é zero.

TEOREMA 1.5. *Se $S = \{x_1, x_2, \dots, x_k\}$ é um conjunto independente formado por k elementos de um espaço linear V e se $L(S)$ é o subespaço gerado por S , então todo o conjunto de $k + 1$ elementos de $L(S)$ é dependente.*

Demonstração. A demonstração faz-se por indução em k , que representa o número de elementos de S . Em primeiro lugar suponhamos $k = 1$. Então, por hipótese, S é formado por um único elemento x_1 , com $x_1 \neq 0$, visto que S é independente. Consideremos agora dois quaisquer elementos distintos y_1 e y_2 em $L(S)$. Então cada um destes elementos é um escalar multiplicado por x_1 , seja $y_1 = c_1 x_1$ e $y_2 = c_2 x_1$, onde c_1 e c_2 são ambos diferentes de 0. Multiplicando y_1 por c_2 e y_2 por c_1 e subtraindo, obtemos

$$c_2 y_1 - c_1 y_2 = 0.$$

Esta é uma representação não trivial de 0, pelo que y_1 e y_2 serão dependentes; está, pois, demonstrado o teorema quando $k = 1$.

Admitamos agora que o teorema é verdadeiro para $k - 1$ e provemos que é ainda verdadeiro para k . Tomemos um conjunto de $k + 1$ elementos em $L(S)$, digamos $T = \{y_1, y_2, \dots, y_{k+1}\}$. Desejamos provar que T é dependente. Visto que cada y_i está em $L(S)$ podemos escrever

$$(1.4) \quad y_i = \sum_{j=1}^k a_{ij} x_j$$

para cada $i = 1, 2, \dots, k + 1$. Examinemos todos os escalares a_{i1} que multiplicam x_1 e para tal devidamos a demonstração em duas partes conforme todos estes escalares são ou não nulos.

CASO 1. $a_{i1} = 0$ para cada $i = 1, 2, \dots, k + 1$. Neste caso a soma em (1.4) não contém x_1 , pelo que cada y_i em T está no subespaço linear gerado pelo conjunto $S' = \{x_2, \dots, x_k\}$. Mas S' é independente e consta de $k - 1$ elementos. Pela hipótese

de indução, o teorema é verdadeiro para $k-1$ pelo que o conjunto T é dependente. Está assim demonstrado o teorema no Caso 1.

CASO 2. Nem todos os escalares a_{i1} são nulos. Admitamos que $a_{11} \neq 0$. (Se necessário, podemos numerar de novo os y de modo a que isso se verifique.) Fazendo $i=1$ na equação (1.4) e multiplicando ambos os membros por c_i , com $c_i = a_{i1}/a_{11}$, obtemos

$$c_i y_1 = a_{i1} x_1 + \sum_{j=2}^k c_i a_{1j} x_j.$$

Se a esta subtrairmos, membro a membro, a equação (1.4) resulta

$$c_i y_1 - y_i = \sum_{j=2}^k (c_i a_{1j} - a_{ij}) x_j,$$

para $i = 2, \dots, k+1$. Esta equação exprime cada um dos k elementos $c_i y_1 - y_i$ como uma combinação linear de $k-1$ elementos independentes x_2, \dots, x_k . Pela hipótese de indução, os k elementos $c_i y_1 - y_i$ devem ser dependentes. Consequentemente para determinada escolha dos escalares t_2, \dots, t_{k+1} não simultaneamente nulos, temos:

$$\sum_{i=2}^{k+1} t_i (c_i y_1 - y_i) = 0,$$

donde resulta

$$\left(\sum_{i=2}^{k+1} t_i c_i \right) y_1 - \sum_{i=2}^{k+1} t_i y_i = 0.$$

Esta, porém, é uma combinação linear não trivial de y_1, \dots, y_{k+1} que representa o elemento zero, pelo que os elementos y_1, \dots, y_{k+1} devem ser dependentes, ficando assim completado a demonstração.

1.8. Bases e dimensão

DEFINIÇÃO. Um conjunto finito S de elementos num espaço linear V chama-se uma base finita de V se S é independente e gera V . O espaço V diz-se de dimensão finita se têm uma base finita, ou se V é formado unicamente por O . Caso contrário V diz-se de dimensão infinita.

TEOREMA 1.6. Se V é um espaço linear de dimensão finita, então cada base finita de V tem o mesmo número de elementos.

Demonstração. Sejam S e T duas bases finitas de V . Suponhamos que S é formado por k elementos e T é formada por m elementos. Uma vez que S é independente e gera V , o teorema 1.5 diz-nos que cada conjunto de $k+1$ elementos de V é dependente. Por conseguinte, todo o conjunto de mais do que k elementos de V é dependente. Visto

que T é um conjunto independente, devemos ter $m \leq k$. O mesmo raciocínio com S e T permutados mostra que $k \leq m$. Portanto $k = m$.

DEFINIÇÃO. Se um espaço linear V tem uma base com n elementos, o inteiro n chama-se *dimensão* de V , e escreve-se $n = \dim V$. Se $V = \{0\}$ diz-se que V tem *dimensão* 0.

EXEMPLO 1. O espaço V_n tem dimensão n . Uma base deste espaço é o conjunto dos n vectores coordenados unitários.

EXEMPLO 2. O espaço de todos os polinómios $p(t)$ de grau $\leq n$ tem dimensão $n + 1$. Uma base é o conjunto de $n + 1$ polinómios $\{1, t, t^2, \dots, t^n\}$. Todo o polinómio de grau $\leq n$ é uma combinação linear desses $n + 1$ polinómios.

EXEMPLO 3. O espaço das soluções da equação diferencial $y'' - 2y' - 3y = 0$ tem dimensão 2. Uma base consiste das duas funções $u_1(x) = e^{-x}$, $u_2(x) = e^{3x}$. Toda a solução é uma combinação linear destas duas.

EXEMPLO 4. O espaço de todos os polinómios $p(t)$ é de dimensão infinita. O conjunto infinito $\{1, t, t^2, \dots\}$ gera este espaço e nenhum conjunto *finito* de polinómios gera o espaço.

TEOREMA 1.7 *Seja V um espaço linear de dimensão finita com $\dim V = n$. Então verifica-se que:*

- (a) *Todo o conjunto de elementos independentes de V é um subconjunto de alguma base de V .*
- (b) *Todo o conjunto de n elementos independentes é uma base para V .*

Demonstração. Para demonstrar (a), designamos por $S = \{x_1, \dots, x_k\}$ qualquer conjunto independente de elementos de V . Se $L(S) = V$, então S é uma base. Caso contrário, existe algum elemento y em V , o qual não pertencerá a $L(S)$. Juntemos este elemento a S e consideremos o novo conjunto $S' = \{x_1, \dots, x_k, y\}$. Se este conjunto fosse dependente existiriam escalares c_1, \dots, c_{k+1} , não todos nulos, tais que

$$\sum_{i=1}^k c_i x_i + c_{k+1} y = 0.$$

Mas $c_{k+1} \neq 0$ visto x_1, \dots, x_k serem independentes. Consequentemente, podemos resolver esta equação em relação a y , chegando à conclusão de que $y \in L(S)$, o que contradiz o facto de que y não pertence a $L(S)$. Portanto, o conjunto S' é independente e contém $k + 1$ elementos. Se $L(S') = V$, então S' é uma base e, visto ser S um subconjunto de S' , a alínea (a) está demonstrada. Se S' não é uma base, podemos raciocinar de novo com S' como o fizemos com S , obtendo um novo conjunto S'' o qual conterá $k + 2$ elementos e será independente. Se S'' é uma base, então a alínea (a) está demonstrada. Caso contrário, repete-se o processo. Devemos assim chegar a uma base ao fim de um número finito de etapas, doutro modo obteríamos eventualmente um conjunto independente com $n + 1$ elementos, contradizendo o teorema 1.5. Por isso, a alínea (a) do teorema 1.7 está demonstrada.

Para demonstrar a alínea (b), designemos por S qualquer conjunto independente formado por n elementos. Devido à alínea (a), S é um subconjunto de certa base, por exemplo B . Mas pelo teorema 1.6, a base B tem precisamente n elementos, e assim $S = B$.

1.9. Componentes

Seja V um espaço linear de dimensão n e consideremos uma base cujos elementos e_1, \dots, e_n se tomam segundo determinada ordem. Representamos uma tal base ordenada por um n -sistema (e_1, \dots, e_n) . Se $x \in V$, podemos exprimir x como uma combinação linear destes elementos da base:

$$(1.5) \quad x = \sum_{i=1}^n c_i e_i.$$

Os coeficientes nesta igualdade determinam um n -sistema de números (c_1, c_2, \dots, c_n) o qual fica univocamente determinado para x . Com efeito, se tivéssemos outra representação de x como combinação linear de e_1, \dots, e_n por exemplo $x = \sum_{i=1}^n d_i e_i$, então subtraindo membro a membro de (1.5), encontramos $\sum_{i=1}^n (c_i - d_i) e_i = 0$. Mas porque os elementos de base são independentes, a igualdade anterior implica $c_i = d_i$ para todo o $i = 1, 2, \dots, n$, pelo que será $(c_1, c_2, \dots, c_n) = (d_1, d_2, \dots, d_n)$.

Os elementos do n -sistema ordenado (c_1, c_2, \dots, c_n) definidos por (1.5) dizem-se *as componentes de x relativamente à base ordenada (e_1, e_2, \dots, e_n)* .

1.10. Exercícios

Em cada um dos Exercícios 1 a 10, S é o conjunto de todos os vectores (x, y, z) de V_3 cujas componentes satisfazem à condição dada. Determinar se S é um subespaço de V_3 . Se S for um subespaço, calcular $\dim S$.

- | | |
|----------------------|---|
| 1. $x = 0$. | 6. $x = y$ ou $x = z$. |
| 2. $x + y = 0$. | 7. $x^2 - y^2 = 0$. |
| 3. $x + y + z = 0$. | 8. $x + y = 1$. |
| 4. $x = y$. | 9. $y = 2x$ e $z = 3x$. |
| 5. $x = y = z$. | 10. $x + y + z = 0$ e $x - y - z = 0$. |

Seja P_n o espaço linear de todos os polinómios de grau $\leq n$, com n fixo. Em cada um dos Exercícios 11 a 20, represente S o conjunto de todos os polinómios f em P_n satisfazendo às condições dadas. Determinar se sim ou não S é um subespaço de P_n . Se S for um subespaço, calcular $\dim S$.

- | | |
|--------------------------|--|
| 11. $f(0) = 0$. | 16. $f(0) = f(2)$. |
| 12. $f'(0) = 0$. | 17. f é par. |
| 13. $f''(0) = 0$. | 18. f é ímpar. |
| 14. $f(0) + f'(0) = 0$. | 19. f tem grau $\leq k$, com $k < n$, ou $f = 0$. |
| 15. $f(0) = f(1)$. | 20. f tem grau k , com $k < n$, ou $f = 0$. |

21. No espaço linear de todos os polinómios reais $p(t)$, descrever o subespaço gerado por cada um dos seguintes subconjuntos de polinómios e determinar a dimensão desse subespaço.
 (a) $\{1, t^2, t^4\}$; (b) $\{t, t^3, t^5\}$; (c) $\{t, t^2\}$; (d) $\{1 + t, (1 + t)^2\}$.
22. Neste Exercício, $L(S)$ representa o subespaço gerado por um subconjunto S de um espaço linear V . Provar cada uma das proposições seguintes de (a) a (f).
 (a) $S \subseteq L(S)$.
 (b) Se $S \subseteq T \subseteq V$ e se T é um subespaço de V , então $L(S) \subseteq T$. Esta propriedade enuncia-se dizendo que $L(S)$ é o *menor* subespaço de V que contém S .
 (c) Um subconjunto S de V é um subespaço de V se e só se $L(S) = S$.
 (d) Se $S \subseteq T \subseteq V$, então $L(S) \subseteq L(T)$.
 (e) Se S e T são subespaços de V , então também o é $S \cap T$.
 (f) Se S e T são subconjuntos de V , então $L(S \cap T) \subseteq L(S) \cap L(T)$.
 (g) Dar um exemplo no qual $L(S \cap T) \neq L(S) \cap L(T)$.
23. Seja V o espaço linear de todas as funções reais definidas na recta real. Determinar se cada um dos seguintes subconjuntos de V é dependente ou independente. Calcular a dimensão do subespaço gerado por cada conjunto.
 (a) $\{1, e^{ax}, e^{bx}\}$, $a \neq b$. (f) $\{\cos x, \sin x\}$.
 (b) $\{e^{ax}, xe^{ax}\}$. (g) $\{\cos^2 x, \sin^2 x\}$.
 (c) $\{1, e^{ax}, xe^{ax}\}$. (h) $\{1, \cos 2x, \sin^2 x\}$.
 (d) $\{e^{ax}, xe^{ax}, x^2 e^{ax}\}$. (i) $\{\sin x, \sin 2x\}$.
 (e) $\{e^x, e^{-x}, \chi x\}$. (j) $\{e^x \cos x, e^{-x} \sin x\}$.
24. Seja V um espaço linear de dimensão finita e S um subespaço de V . Demonstrar cada uma das seguintes proposições.
 (a) S tem dimensão finita e $\dim S \leq \dim V$.
 (b) $\dim S = \dim V$ se e só se $S = V$.
 (c) Toda a base de S é parte de uma base de V .
 (d) Uma base de V não contém necessariamente uma base de S .

1.11. Produto interno, espaços euclidianos. Normas

Na geometria euclidiana, aquelas propriedades que contam com a possibilidade de medição de comprimento de segmentos de recta e ângulos definidos por rectas chamam-se propriedades *métricas*. No nosso estudo de V_n , definimos comprimento e ângulos a partir do produto escalar. Desejamos agora generalizar aquelas noções a espaços lineares mais gerais. Introduziremos em primeiro lugar uma generalização do produto escalar que designaremos agora por *produto interno* e definiremos comprimentos e ângulos em função do produto interno.

O produto escalar $x \cdot y$ de dois vectores $x = (x_1, x_2, \dots, x_n)$ e $y = (y_1, y_2, \dots, y_n)$ em V_n foi definido, no Capítulo 12, pela fórmula (1.6).

$$(1.6) \quad x \cdot y = \sum_{i=1}^n x_i y_i.$$

Num espaço linear qualquer, escrevemos (x, y) , em vez de $x \cdot y$, para o produto interno e definimos este axiomáticamente, em vez de o fazermos por uma fórmula específica, isto é, estabelecemos um certo número de propriedades que pretendemos que o produto interno possua e consideramo-las como *axiomas*.

DEFINIÇÃO. Define-se num espaço linear real V um produto interno se a cada par de elementos x e y em V corresponde um único número real (x, y) satisfazendo aos seguintes axiomas, quaisquer que sejam x, y e z de V e qualquer que seja o escalar real c .

- (1) $(x, y) = (y, x)$ (comutatividade, ou simetria).
- (2) $(x, y + z) = (x, y) + (x, z)$ (distributividade, ou linearidade).
- (3) $c(x, y) = (cx, y)$ (associatividade, ou homogeneidade).
- (4) $(x, x) > 0$ se $x \neq 0$ (positividade).

Um espaço linear dotado com a operação produto interno diz-se um *espaço euclidiano real*.

Nota. Fazendo $c = 0$ em (3), encontramos $(0, y) = 0$ para todo o y .

Num espaço linear complexo, um produto interno (x, y) é um número complexo satisfazendo aos mesmos axiomas que os do produto interno real, excepto o axioma de simetria que é substituído pela relação

$$(1') \quad (x, y) = \overline{(y, x)}, \quad (\text{simetria hermitica}^\dagger)$$

onde $\overline{(y, x)}$ representa o complexo conjugado de (y, x) . No axioma da homogeneidade, o factor escalar c pode ser qualquer número complexo. Do axioma da homogeneidade e (1') obtemos a relação

$$(3') \quad (x, cy) = \overline{(cy, x)} = \overline{c(y, x)} = \bar{c}(x, y).$$

Um espaço linear complexo dotado de produto interno chama-se um *espaço euclidiano complexo*. (Algumas vezes também se usa a designação *espaço unitário*). Um exemplo é o espaço vectorial complexo $V_n(\mathbb{C})$ já referido na Secção 12.16 do Volume I.

Embora o nosso interesse resida fundamentalmente nos exemplos de espaços euclidianos reais, os teoremas que se apresentam a seguir são válidos igualmente para espaços euclidianos complexos. Quando usarmos a expressão *espaço euclidiano*, sem fazer qualquer referência complementar, subentender-se-á que o espaço pode ser real ou complexo.

O leitor verificará que cada um dos exemplos seguintes satisfaz a todos os axiomas do produto interno.

EXEMPLO 1. Em V_n seja $(x, y) = x \cdot y$, o produto escalar usual de x por y .

EXEMPLO 2. Se $x = (x_1, x_2)$ e $y = (y_1, y_2)$ são dois vectores quaisquer de V_2 , definir (x, y) pela fórmula

$$(x, y) = 2x_1y_1 + x_1y_2 + x_2y_1 + x_2y_2.$$

(†) Em honra de Charles Hermite (1822-1901), um matemático francês que contribuiu muito para o desenvolvimento da álgebra e análise.

Este exemplo mostra que pode estar definido mais do que um produto interno num dado espaço linear.

EXEMPLO 3. Represente $C(a, b)$ o espaço linear de todas as funções reais contínuas definidas num intervalo $[a, b]$. Definamos o produto interno de duas funções f e g pela fórmula

$$(f, g) = \int_a^b f(t)g(t) dt.$$

Esta fórmula é análoga à Equação (1.6) que define o produto escalar de dois vectores de V_n . Os valores das funções $f(t)$ e $g(t)$ desempenham o papel das componentes x_i e y_i e a integração substitui a somação.

EXEMPLO 4. No espaço $C(a, b)$, definimos

$$(f, g) = \int_a^b w(t)f(t)g(t) dt,$$

com w uma função positiva dada em $C(a, b)$. A função w diz-se a *função peso*. No exemplo 3 temos $w(t) = 1$ para todo o t .

EXEMPLO 5. No espaço linear dos polinómios reais, definimos

$$(f, g) = \int_0^\infty e^{-t}f(t)g(t) dt.$$

Em virtude do factor exponencial, este integral impróprio converge para todo o par de polinómios f e g .

TEOREMA 1.8. *Num espaço euclidiano V , todo o produto interno verifica a desigualdade de Cauchy-Schwarz:*

$$|(x, y)|^2 \leq (x, x)(y, y) \quad \text{quaisquer que sejam } x \text{ e } y \text{ em } V.$$

Além disso, o sinal de igualdade verifica-se se e só se x e y são dependentes.

Demonstração. Se acontece que ou $x = 0$ ou $y = 0$ a demonstração é trivial, pelo que supomos que ambos x e y são não nulos. Seja $z = ax + by$, com a e b escalares a serem especificados mais adiante. Temos a desigualdade $(z, z) \geq 0$ para todo o par a e b . Quando explicitamos esta desigualdade em função de x e y com uma escolha adequada de a e b , obtemos a desigualdade de Cauchy-Schwarz.

Para exprimir (z, z) em função de x e y servimo-nos das propriedades (1'), (2) e (3') e concluímos que

$$\begin{aligned}(z, z) &= (ax + by, ax + by) = (ax, ax) + (ax, by) + (by, ax) + (by, by) \\ &= a\bar{a}(x, x) + a\bar{b}(x, y) + b\bar{a}(y, x) + b\bar{b}(y, y) \geq 0.\end{aligned}$$

Tomando $a = (y, y)$ e suprimindo na desigualdade o factor positivo (y, y) resulta

$$(y, y)(x, x) + \bar{b}(x, y) + b(y, x) + b\bar{b} \geq 0.$$

Façamos agora $b = -(x, y)$. Então $\bar{b} = -(y, x)$ e a última desigualdade simplifica-se, tomando a forma

$$(y, y)(x, x) \geq (x, y)(y, x) = |(x, y)|^2.$$

o que prova a desigualdade de Cauchy-Schwarz. O sinal de igualdade é válido através da demonstração, se e só se $z = 0$. Isto verifica-se, em particular, se e só se x e y são dependentes.

EXEMPLO. Aplicando a teorema 1.8 ao espaço $C(a, b)$ com o produto interno $(f, g) = \int_a^b f(t)g(t)dt$, encontramos para a desigualdade de Cauchy-Schwarz

$$\left(\int_a^b f(t)g(t) dt \right)^2 \leq \left(\int_a^b f^2(t) dt \right) \left(\int_a^b g^2(t) dt \right).$$

O produto interno pode ser usado para introduzir o conceito métrico de comprimento em qualquer espaço euclidiano.

DEFINIÇÃO. Num espaço euclidiano V , o número não negativo $\|x\|$ definido pela igualdade

$$\|x\| = (x, x)^{1/2}$$

chama-se a norma do elemento x .

Expressando a desigualdade de Cauchy-Schwarz em termos de normas escreve-se

$$|(x, y)| \leq \|x\| \|y\|.$$

Visto ser possível definir um produto interno de diferentes maneiras, a norma de um elemento dependerá da escolha do produto interno. Esta falta de unicidade era de esperar. Tal facto é análogo ao de podermos atribuir diferentes números à medida do comprimento de dado segmento de recta, dependendo da escolha da unidade de medida. O teorema seguinte define propriedades fundamentais das normas que não dependem da escolha do produto interno.

TEOREMA 1.9. *Num espaço euclidiano, toda a norma goza das seguintes propriedades para todos os elementos x e y , e todo o escalar c :*

- (a) $\|x\| = 0$ se $x = O$.
- (b) $\|x\| > 0$ se $x \neq O$, (positividade).
- (c) $\|cx\| = |c|\|x\|$ (homogeneidade).
- (d) $\|x + y\| \leq \|x\| + \|y\|$ (desigualdade triangular).

O sinal de igualdade verifica-se em (d) se $x = O$, se $y = O$, ou se $y = cx$ para algum $c > 0$.

Demonstração. As propriedades (a), (b) e (c) deduzem-se imediatamente dos axiomas do produto interno. Para demonstrar (d), observemos que

$$\begin{aligned}\|x + y\|^2 &= (x + y, x + y) = (x, x) + (y, y) + (x, y) + (y, x) \\ &= \|x\|^2 + \|y\|^2 + (x, y) + \overline{(x, y)}.\end{aligned}$$

A soma $(x, y) + \overline{(x, y)}$ é real. A desigualdade de Cauchy-Schwarz mostra que $|(x, y)| \leq \|x\| \|y\|$ e $|\overline{(x, y)}| \leq \|x\| \|y\|$, pelo que se tem

$$\|x + y\|^2 \leq \|x\|^2 + \|y\|^2 + 2\|x\| \|y\| = (\|x\| + \|y\|)^2.$$

o que demonstra (d). Quando $y = cx$, com $c > 0$, temos

$$\|x + y\| = \|x + cx\| = (1 + c)\|x\| = \|x\| + \|cx\| = \|x\| + \|y\|.$$

DEFINIÇÃO. *Num espaço euclidiano real V , o ângulo definido por dois elementos não nulos x e y define-se como sendo o número θ do intervalo $0 \leq \theta \leq \pi$ dado por*

$$(1.7) \quad \cos \theta = \frac{(x, y)}{\|x\| \|y\|}.$$

Nota: A desigualdade de Cauchy-Schwarz mostra que o valor do quociente no segundo membro de (1.7) pertence ao intervalo $[-1, 1]$, pelo que existe um e um só θ no intervalo $[0, \pi]$ cujo cosseno é igual ao valor daquele quociente.

1.12. Ortogonalidade num espaço euclidiano

DEFINIÇÃO. *Num espaço euclidiano V , dois elementos x e y dizem-se ortogonais se o correspondente produto interno for zero. Um subconjunto S de V diz-se um subconjunto ortogonal se $(x, y) = 0$ para cada par de elementos distintos x e y de S . Um conjunto ortogonal diz-se ortonormado se cada um dos seus elementos tem norma 1.*

O elemento zero é ortogonal a todo o elemento de V ; é o único elemento ortogonal a si próprio. O teorema seguinte mostra uma relação entre ortogonalidade e dependência.

TEOREMA 1.10. *Num espaço euclidiano V , todo o conjunto ortogonal de elementos não nulos é independente. Em particular, num espaço euclidiano de dimensão finita com $\dim V = n$, todo o conjunto ortogonal formado por n elementos não nulos define uma base de V .*

Demonstração. Seja S um conjunto ortogonal de elementos não nulos de V , e suponhamos que certa combinação linear finita de elementos de S é igual a zero, quer dizer

$$\sum_{i=1}^k c_i x_i = 0,$$

onde cada $x_i \in S$. Multiplicando escalarmente ambos os membros por x_1 e tendo presente que $(x_i, x_i) = 0$ se $i \neq 1$, encontramos que $c_1(x_1, x_1) = 0$. Mas $(x_1, x_1) \neq 0$ visto que $x_1 \neq 0$, donde resulta $c_1 = 0$. Repetindo o raciocínio com x_1 substituído por x_j , encontramos cada $c_j = 0$, o que prova que S é independente. Se $\dim V = n$ e se S é formado por n elementos, o teorema 1.7(b) mostra que S é uma base de V .

EXEMPLO. No espaço linear real $C(0, 2\pi)$ com o produto interno $(f, g) = \int_0^{2\pi} f(x)g(x) dx$, seja S o conjunto de funções trigonométricas (u_0, u_1, \dots) definido da seguinte maneira

$$u_0(x) = 1, \quad u_{2n-1}(x) = \cos nx, \quad u_{2n}(x) = \sin nx, \quad \text{para } n = 1, 2, \dots$$

Se $m \neq n$, temos as relações de ortogonalidade

$$\int_0^{2\pi} u_n(x)u_m(x) dx = 0,$$

e portanto S é um conjunto ortogonal. Visto que nenhum elemento de S é o elemento zero, S é independente. A norma de cada elemento de S calcula-se facilmente. Temos $(u_0, u_0) = \int_0^{2\pi} dx = 2\pi$ e, para $n \geq 1$, temos

$$(u_{2n-1}, u_{2n-1}) = \int_0^{2\pi} \cos^2 nx dx = \pi, \quad (u_{2n}, u_{2n}) = \int_0^{2\pi} \sin^2 nx dx = \pi.$$

Por conseguinte, $\|u_0\| = \sqrt{2\pi}$ e $\|u_n\| = \sqrt{\pi}$ para $n \geq 1$. Dividindo cada u_n pela respectiva norma, obtemos um conjunto ortonormado $\{\varphi_0, \varphi_1, \varphi_2, \dots\}$ com $\varphi_n = u_n/\|u_n\|$. Então resulta

$$\varphi_0(x) = \frac{1}{\sqrt{2\pi}}, \quad \varphi_{2n-1}(x) = \frac{\cos nx}{\sqrt{\pi}}, \quad \varphi_{2n}(x) = \frac{\sin nx}{\sqrt{\pi}}, \quad \text{para } n \geq 1.$$

Na Secção 1.14 provaremos que todo o espaço euclidiano de dimensão finita admite uma base ortogonal. O teorema que se segue mostra precisamente como calcular as componentes de um elemento relativamente a tal base.

TEOREMA 1.11. *Seja V um espaço euclidiano de dimensão n , finita, e admita-se que $S = \{e_1, e_2, \dots, e_n\}$ é uma base ortogonal de V . Se um elemento x se exprime como uma combinação linear dos elementos da base, seja*

$$(1.8) \quad x = \sum_{i=1}^n c_i e_i,$$

então as suas componentes na base ordenada (e_1, e_2, \dots, e_n) são dadas pelas fórmulas

$$(1.9) \quad c_j = \frac{(x, e_j)}{(e_j, e_j)} \quad \text{para } j = 1, 2, \dots, n.$$

Em particular, se S admite uma base ortonormada, cada c_j é dado por

$$(1.10) \quad c_j = (x, e_j).$$

Demonstração. Efectuando o produto interno de cada membro de (1.8) por e_j , obtemos

$$(x, e_j) = \sum_{i=1}^n c_i (e_i, e_j) = c_j (e_j, e_j)$$

visto que $(e_i, e_j) = 0$ se $i \neq j$. Isto implica (1.9), e quando $(e_j, e_j) = 1$, obtemos (1.10).

Se $\{e_1, e_2, \dots, e_n\}$ é uma base ortonormada, a equação (1.9) pode escrever-se na forma

$$(1.11) \quad x = \sum_{i=1}^n (x, e_i) e_i.$$

O teorema seguinte mostra que, num espaço euclidiano com dimensão finita com uma base ortonormada, o produto interno de dois elementos pode ser calculado em função das respectivas componentes.

TEOREMA 1.12. *Seja V um espaço euclidiano de dimensão finita n , e admita-se que $\{e_1, e_2, \dots, e_n\}$ é uma base ortonormada de V . Então para todo o par de elementos x e y de V , tem-se*

$$(1.12) \quad (x, y) = \sum_{i=1}^n (x, e_i) \overline{(y, e_i)} \quad (\text{fórmula de Parseval})$$

Em particular, quando $x = y$, tem-se

$$(1.13) \quad \|x\|^2 = \sum_{i=1}^n |(x, e_i)|^2.$$

Demonstração. Multiplicando internamente ambos os membros da equação (1.11) por y e aplicando a propriedade linearidade do produto interno, obtemos (1.12). Quando $x = y$, a equação (1.12) reduz-se a (1.13).

Nota: A fórmula (1.12) tem a designação que se indica, em honra de M.A. Parseval (aproximadamente 1776-1836) que obteve este tipo de fórmula num espaço funcional especial. A equação (1.13) é uma generalização do teorema de Pitágoras.

1.13. Exercícios

- Sejam $x = (x_1, \dots, x_n)$ e $y = (y_1, \dots, y_n)$ vectores arbitrários de V_n . Para cada alínea, determinar se (x, y) , definido pela fórmula que se indica, é um produto interno para V_n . Caso (x, y) não seja um produto interno, dizer quais os axiomas que não são verificados.

$$(a) \quad (x, y) = \sum_{i=1}^n x_i |y_i|.$$

$$(d) \quad (x, y) = \left(\sum_{i=1}^n x_i^2 y_i^2 \right)^{1/2}.$$

$$(b) \quad (x, y) = \left| \sum_{i=1}^n x_i y_i \right|.$$

$$(e) \quad (x, y) = \sum_{i=1}^n (x_i + y_i)^2 - \sum_{i=1}^n x_i^2 - \sum_{i=1}^n y_i^2.$$

$$(c) \quad (x, y) = \sum_{i=1}^n x_i \sum_{j=1}^n y_j.$$

- Suponhamos que retemos os três primeiros axiomas para o produto interno real (simetria, linearidades e homogeneidade), mas que substituímos o quatro axioma por um novo axioma (4'): $(x, x) = 0$ se e só se $x = 0$. Provar que será para todo o $x \neq 0$ ou $(x, x) > 0$, ou $(x, x) < 0$.

[Sugestão: Supor $(x, x) > 0$ para certo $x \neq 0$ e $(y, y) < 0$ para certo $y \neq 0$. No espaço gerado por $\{x, y\}$, determinar um elemento $z \neq 0$ com $(z, z) = 0$].

Provar que cada uma das proposições dos Exercícios 3 a 7 é válida para todos os pares de elementos x e y de um espaço euclídeo real.

- $(x, y) = 0$ se e só se $\|x + y\| = \|x - y\|$.
- $(x, y) = 0$ se e só se $\|x + y\|^2 = \|x\|^2 + \|y\|^2$.
- $(x, y) = 0$ se e só se $\|x + cy\| \geq \|x\|$ para todo o real c .
- $(x + y, x - y) = 0$ se e só se $\|x\| = \|y\|$.
- Se x e y são elementos não nulos, definindo entre si um ângulo θ , então

$$\|x - y\|^2 = \|x\|^2 + \|y\|^2 - 2 \|x\| \|y\| \cos \theta.$$

8. No espaço linear real $C(1, e)$, definir um produto interno por

$$(f, g) = \int_1^e (\log x) f(x) g(x) dx.$$

- (a) Se $f(x) = \sqrt{x}$, calcular $\|f\|$.
 (b) Determinar um polinómio linear $g(x) = a + bx$ que seja ortogonal à função constante $f(x) = 1$.
9. No espaço linear real $C(-1, 1)$ seja $(f, g) = \int_{-1}^1 f(t)g(t) dt$. Considerar as três funções u_1, u_2, u_3 dadas por

$$u_1(t) = 1, \quad u_2(t) = t, \quad u_3(t) = 1 + t.$$

Provar que duas delas são ortogonais, duas definem um ângulo de $\pi/3$ e duas outras definem um ângulo de $\pi/6$.

10. No espaço linear P_n de todos os polinómios reais de grau $\leq n$, definir

$$(f, g) = \sum_{k=0}^n f\left(\frac{k}{n}\right) g\left(\frac{k}{n}\right).$$

- (a) Provar que (f, g) é um produto interno para P_n .
 (b) Calcular (f, g) quando $f(t) = t$ e $g(t) = at + b$.
 (c) Se $f(t) = t$, determinar todos os polinómios lineares g ortogonais a f .
11. No espaço linear de todos os polinómios reais, definir $(f, g) = \int_0^{\infty} e^{-t} f(t)g(t) dt$.
 (a) Provar que este integral impróprio converge absolutamente para quaisquer polinómios f e g .
 (b) Se $x_n(t) = t^n$ para $n = 0, 1, 2, \dots$, provar que $(x_n, x_m) = (m + n)!$.
 (c) Calcular (f, g) quando $f(t) = (t + 1)^2$ e $g(t) = t^2 + 1$.
 (d) Determinar todos os polinómios lineares $g(t) = a + bt$ ortogonais a $f(t) = 1 + t$.
12. No espaço linear de todos os polinómios reais, determinar se sim ou não (f, g) é um produto interno quando (f, g) é definido pela fórmula indicada. Caso (f, g) não seja um produto interno, indicar quais os axiomas que não são verificados. Em (c), f' e g' representam derivadas.

$$(a) (f, g) = f(1)g(1).$$

$$(c) (f, g) = \int_0^1 f'(t)g'(t) dt.$$

$$(b) (f, g) = \left| \int_0^1 f(t)g(t) dt \right|.$$

$$(d) (f, g) = \left(\int_0^1 f(t) dt \right) \left(\int_0^1 g(t) dt \right).$$

13. V é formado por todas as sucessões infinitas $\{x_n\}$ de números reais para as quais as séries $\sum x_n^2$ convergem. Se $x = \{x_n\}$ e $y = \{y_n\}$ são dois elementos de V , define-se

$$(x, y) = \sum_{n=1}^{\infty} x_n y_n.$$

- (a) Provar que esta série converge absolutamente:

[Sugestão. Utilizar a desigualdade de Cauchy-Schwarz para estimar a soma $\sum_{n=1}^M |x_n y_n|$.]

- (b) Provar que V é um espaço linear com (x, y) como produto interno.
 (c) Calcular (x, y) se $x_n = 1/n$ e $y_n = 1/(n+1)$ para $n \geq 1$.
 (d) Calcular (x, y) se $x_n = 2^n$ e $y_n = 1/n!$ para $n \geq 1$.
14. Seja V o conjunto de todas as funções reais f contínuas em $[0, +\infty)$ e tais que o integral $\int_0^\infty e^{-t} f^2(t) dt$ converge. Definir $(f, g) = \int_0^\infty e^{-t} f(t)g(t) dt$.
 (a) Provar que o integral para (f, g) converge absolutamente para cada par de funções f e g em V .
 [Sugestão: Utilizar a desigualdade de Cauchy-Schwarz para estimar o integral $\int_0^\infty e^{-t} |f(t)g(t)| dt$.]
 (b) Provar que V é um espaço linear com (f, g) como produto interno.
 (c) Calcular (f, g) se $f(t) = e^{-t}$ e $g(t) = t^n$, com $n = 0, 1, 2, \dots$.
15. Num espaço euclidiano complexo, provar que o produto interno tem as seguintes propriedades para todos os elementos x, y e z e todos os complexos a e b .
 (a) $(ax, by) = \bar{a}\bar{b}(x, y)$ (b) $(x, ay + bz) = \bar{a}(x, y) + \bar{b}(x, z)$.
16. Provar que as identidades seguintes são válidas em todo o espaço euclidiano.
 (a) $\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2(x, y)$
 (b) $\|x + y\|^2 - \|x - y\|^2 = 4(x, y)$
 (c) $\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2$.
17. Provar que o espaço de todas as funções complexas contínuas num intervalo $[a, b]$ se transforma num espaço unitário se definirmos um produto interno pela fórmula

$$(f, g) = \int_a^b w(t) f(t) \overline{g(t)} dt,$$

com w uma função positiva dada, contínua em $[a, b]$.

1.14 Construção de conjuntos ortogonais. O método de Gram-Schmidt

Todo o espaço linear de dimensão finita possui uma base finita. Se o espaço é euclidiano, podemos sempre construir uma base *ortogonal*. Este resultado será deduzido como uma consequência de um teorema geral, cuja demonstração ensina a construir conjuntos ortogonais, em qualquer espaço euclidiano, com dimensão finita ou infinita. A processo de construção chama-se o *método de ortogonalização de Gram-Schmidt*, em honra de J. P. Gram (1850-1916) e E. Schmidt (1845-1921).

TEOREMA 1.13 TEOREMA DE ORTOGONALIZAÇÃO. *Seja x_1, x_2, \dots , uma sucessão finita ou infinita de elementos de um espaço euclidiano V , e seja $L(x_1, \dots, x_k)$ o subespaço gerado pelos primeiros k daqueles elementos. Existe uma sucessão correspondente de elementos y_1, y_2, \dots de V , a qual goza das seguintes propriedades para todo o inteiro k :*

- (a) *O elemento y_k é ortogonal a todo o elemento do subespaço $L(y_1, y_2, \dots, y_{k-1})$.*
 (b) *O subespaço gerado por y_1, \dots, y_k é o mesmo que o gerado por x_1, x_2, \dots, x_k .*

$$L(y_1, \dots, y_k) = L(x_1, \dots, x_k).$$

(c) *A sucessão y_1, y_2, \dots é única, a menos de factores escalares, isto é, se y'_1, y'_2, \dots é outra sucessão de elementos de V satisfazendo às propriedades (a) e (b), então para cada k existe um escalar c_k tal que $y'_k = c_k y_k$.*

Demonstração. Construíamos os elementos y_1, y_2, \dots , por indução. Para iniciar o processo, fazemos $y_1 = x_1$. Suponhamos agora que construímos y_1, \dots, y_r de maneira que (a) e (b) sejam satisfeitas quando $k = r$. Definamos y_{r+1} pela igualdade

$$(1.14) \quad y_{r+1} = x_{r+1} - \sum_{i=1}^r a_i y_i,$$

com os escalares a_1, \dots, a_r a determinar. Para $j \leq r$, o produto interno de y_{r+1} com y_j é dado por

$$(y_{r+1}, y_j) = (x_{r+1}, y_j) - \sum_{i=1}^r a_i (y_i, y_j) = (x_{r+1}, y_j) - a_j (y_j, y_j),$$

visto que $(y_i, y_j) = 0$ se $i \neq j$. Se $y_j \neq 0$, podemos fazer y_{r+1} ortogonal a y_j tomando

$$(1.15) \quad a_j = \frac{(x_{r+1}, y_j)}{(y_j, y_j)}.$$

Se $y_j = 0$, então y_{r+1} é ortogonal a y_j para qualquer escolha de a_j , e neste caso escolhemos $a_j = 0$. Assim, o elemento y_{r+1} fica bem definido e é ortogonal a cada um dos elementos precedentes y_1, \dots, y_r , portanto é ortogonal a todo o elemento do subespaço

$$L(y_1, \dots, y_r).$$

Isto prova (a) quando $k = r + 1$.

Para demonstrar (b) quando $k = r + 1$, devemos provar que $L(y_1, \dots, y_{r+1}) = L(x_1, \dots, x_{r+1})$, dado que $L(y_1, \dots, y_r) = L(x_1, \dots, x_r)$. Os r primeiros elementos y_1, \dots, y_r estão em

$$L(x_1, \dots, x_r)$$

e por isso estão no subespaço mais amplo $L(x_1, \dots, x_{r+1})$. O novo elemento y_{r+1} dado por (1.14) é a diferença de dois elementos de $L(x_1, x_2, \dots, x_{r+1})$, pelo que também está em $L(x_1, \dots, x_{r+1})$. Isto prova que

$$L(y_1, \dots, y_{r+1}) \subseteq L(x_1, \dots, x_{r+1}).$$

A equação (1.14) mostra que x_{r+1} é a soma de dois elementos de $L(y_1, \dots, y_{r+1})$, pelo que um raciocínio semelhante nos conduz à inclusão em sentido inverso:

$$L(x_1, \dots, x_{r+1}) \subseteq L(y_1, \dots, y_{r+1}).$$

ficando, pois, demonstrado (b) quando $k = r + 1$. Assim temos que (a) e (b) foram demonstrados por indução a respeito de k .

Finalmente demonstramos (c) por indução a respeito de k . O caso $k = 1$ é trivial. Deste modo, supomos que (c) é verdadeira para $k = r$ e consideremos o elemento y'_{r+1} . Em virtude de (b), este elemento pertence a

$$L(y_1, \dots, y_{r+1}),$$

pelo que podemos escrever

$$y'_{r+1} = \sum_{i=1}^{r+1} c_i y_i = z_r + c_{r+1} y_{r+1},$$

com $z_r \in L(y_1, \dots, y_r)$. Desejamos provar que $z_r = O$. Pela propriedade (a), quer y'_{r+1} , quer $c_{r+1} y_{r+1}$ são ortogonais a z_r . Deste modo, a sua diferença, z_r , é ortogonal a z_r , isto é, z_r é ortogonal a si próprio; pelo que $z_r = O$. Está, pois, completada a demonstração do teorema de ortogonalização.

Na construção precedente, suponhamos que se tem $y_{r+1} = O$ para algum r . Então (1.14) mostra que x_{r+1} é uma combinação linear de y_1, \dots, y_r , e por isso de x_1, \dots, x_r , pelo que os elementos x_1, \dots, x_{r+1} são dependentes. Por outras palavras, se os primeiros k elementos x_1, \dots, x_k são independentes, então os elementos correspondentes y_1, \dots, y_k são *não nulos*. Neste caso, aos coeficientes a_i em (1.14) são dados por (1.15), e as fórmulas definindo (y_1, \dots, y_k) escrevem-se

$$(1.16) \quad y_1 = x_1, \quad y_{r+1} = x_{r+1} - \sum_{i=1}^r \frac{(x_{r+1}, y_i)}{(y_i, y_i)} y_i \quad \text{para } r = 1, 2, \dots, k-1.$$

Estas fórmulas descrevem o método de Gram-Schmidt para a construção de um conjunto ortogonal de elementos não nulos y_1, \dots, y_k , os quais geram o mesmo sub-espaço que um dado conjunto independente x_1, \dots, x_k . Em particular, se x_1, \dots, x_k é uma base de um espaço euclidiano de dimensão finita, então y_1, \dots, y_k é uma base ortogonal para o mesmo espaço. Podemos ainda converter esta numa base ortonormal pela *normalização* de cada uma dos seus elementos y_i , isto é, pela divisão de cada um pela respectiva norma. Por conseguinte, como um corolário do Teorema 1.13 podemos enunciar

TEOREMA 1.14. *Todo o espaço euclidiano de dimensão finita possui uma base ortonormal.*

Se x e y são elementos de um espaço euclidiano, com $y \neq O$, o elemento

$$\frac{(x, y)}{(y, y)} y$$

diz-se a *projectão de x sobre y* . No método de Gram-Schmidt (1.16), construímos o elemento y_{r+1} subtraindo de x_{r+1} a projectão de x_{r+1} sobre cada um dos anteriores

elementos y_1, \dots, y_r . A figura 1.1 mostra essa construção geométrica no espaço vectorial V_3 .

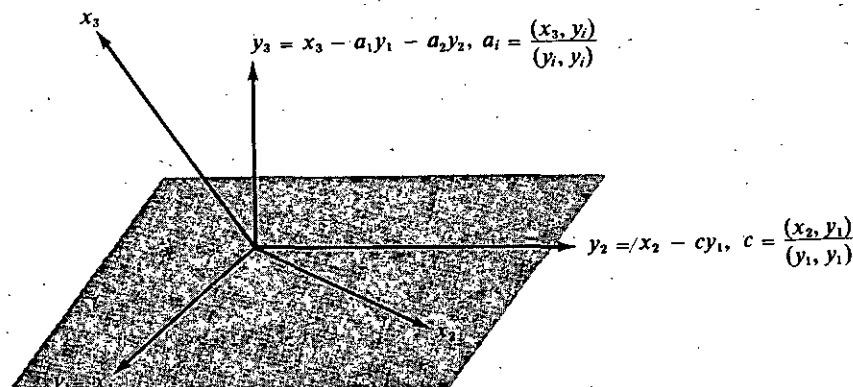


FIG. 1.1. O método de Gram-Schmidt em V_3 . Um conjunto ortogonal $\{y_1, y_2, y_3\}$ foi construído a partir de um conjunto independente dado $\{x_1, x_2, x_3\}$.

EXEMPLO 1. Em V_4 , determinar uma base ortonormada para o subespaço gerado pelos três vectores $x_1 = (1, -1, 1, -1)$, $x_2 = (5, 1, 1, 1)$, e $x_3 = (-3, -3, 1, -3)$.

Resolução. Aplicando o método de Gram-Schmidt, obtemos

$$y_1 = x_1 = (1, -1, 1, -1),$$

$$y_2 = x_2 - \frac{(x_2, y_1)}{(y_1, y_1)} y_1 = x_2 - y_1 = (4, 2, 0, 2),$$

$$y_3 = x_3 - \frac{(x_3, y_1)}{(y_1, y_1)} y_1 - \frac{(x_3, y_2)}{(y_2, y_2)} y_2 = x_3 - y_1 + y_2 = (0, 0, 0, 0).$$

Visto que $y_3 = 0$, os três vectores x_1, x_2, x_3 devem ser dependentes. Mas uma vez que y_1 e y_2 são não nulos, os vectores x_1 e x_2 são independentes. Por conseguinte $L(x_1, x_2, x_3)$ é um subespaço de dimensão 2. O conjunto $\{y_1, y_2\}$ é uma base ortogonal para este subespaço. Dividindo cada um dos y_1 e y_2 pela correspondente norma, obtemos uma base ortonormada formada pelos dois vectores

$$\frac{y_1}{\|y_1\|} = \frac{1}{2}(1, -1, 1, -1) \quad \text{e} \quad \frac{y_2}{\|y_2\|} = \frac{1}{\sqrt{6}}(2, 1, 0, 1).$$

EXEMPLO 2. *Polinómios de Legendre.* No espaço linear de todos os polinómios, com o produto interno $(x, y) = \int_{-1}^1 x(t) y(t) dt$, consideremos a sucessão infinita $x_0, x_1, x_2,$

..., onde $x_n(t) = t^n$. Quando se aplica o teorema de ortogonalização a esta sucessão, obtém-se outra sucessão de polinómios y_0, y_1, y_2, \dots , pela primeira vez encontrados pelo matemático francês A.M. Legendre (1752-1833) nos seus trabalhos sobre a teoria do potencial. Os primeiros desses polinómios calculam-se facilmente pelo método de Gram-Schmidt. Em primeiro lugar temos $y_0(t) = x_0(t) = 1$. Uma vez que

$$(y_0, y_0) = \int_{-1}^1 dt = 2 \quad \text{e} \quad (x_1, y_0) = \int_{-1}^1 t dt = 0,$$

encontramos

$$y_1(t) = x_1(t) - \frac{(x_1, y_0)}{(y_0, y_0)} y_0(t) = x_1(t) = t.$$

Utilizamos depois as relações

$$(x_2, y_0) = \int_{-1}^1 t^2 dt = \frac{2}{3}, \quad (x_2, y_1) = \int_{-1}^1 t^3 dt = 0, \quad (y_1, y_1) = \int_{-1}^1 t^2 dt = \frac{2}{3},$$

para obtermos

$$y_2(t) = x_2(t) - \frac{(x_2, y_0)}{(y_0, y_0)} y_0(t) - \frac{(x_2, y_1)}{(y_1, y_1)} y_1(t) = t^2 - \frac{1}{3}.$$

Analogamente, encontramos

$$y_3(t) = t^3 - \frac{3}{5}t, \quad y_4(t) = t^4 - \frac{6}{7}t^2 + \frac{3}{35}, \quad y_5(t) = t^5 - \frac{10}{9}t^3 + \frac{5}{21}t.$$

Voltaremos a encontrar estes polinómios no Capítulo 6, no estudo complementar de equações diferenciais e provaremos que

$$y_n(t) = \frac{n!}{(2n)!} \frac{d^n}{dt^n} (t^2 - 1)^n.$$

Os polinómios P_n dados por

$$P_n(t) = \frac{(2n)!}{2^n(n!)^2} y_n(t) = \frac{1}{2^n n!} \frac{d^n}{dt^n} (t^2 - 1)^n$$

designam-se por *polinómios de Legendre*. Os polinómios na correspondente sucessão ortonormada $\varphi_0, \varphi_1, \varphi_2, \dots$, dados por $\varphi_n = y_n / \|y_n\|$, chamam-se *polinómios de Legendre normalizados*. A partir das expressões para y_0, \dots, y_5 , dadas atrás, encontramos

$$\varphi_0(t) = \sqrt{\frac{1}{2}}, \quad \varphi_1(t) = \sqrt{\frac{3}{2}} t, \quad \varphi_2(t) = \frac{1}{2} \sqrt{\frac{5}{2}} (3t^2 - 1), \quad \varphi_3(t) = \frac{1}{2} \sqrt{\frac{7}{2}} (5t^3 - 3t),$$

$$\varphi_4(t) = \frac{1}{8} \sqrt{\frac{9}{2}} (35t^4 - 30t^2 + 3), \quad \varphi_5(t) = \frac{1}{8} \sqrt{\frac{11}{2}} (63t^5 - 70t^3 + 15t).$$

1.15. Complementos ortogonais. Projectões.

Seja V um espaço euclidiano e S um subespaço de dimensão finita. Pretendemos analisar o seguinte tipo de problema de aproximação. *Dado um elemento x de V , determinar um elemento de S cuja distância a x seja tão pequena quanto possível.* A distância entre os dois elementos x e y define-se pela norma $\|x - y\|$.

Antes de analisar este problema na sua forma geral, consideremos um caso particular, representado na figura 1.2. Aqui V é o espaço vectorial V_3 e S é um subespaço a duas dimensões, um plano passando pela origem. Dado x em V , o problema consiste em encontrar, no plano S , o ponto s o mais próximo possível de x .

Se $x \in S$, então evidentemente que a solução é $s = x$. Se x não pertence a S , então o ponto s mais próximo define-se pelo pé da perpendicular tirada de x para o plano S . Este exemplo simples sugere uma maneira de abordar o problema geral de aproximação e dá origem a discussão que se segue.

DEFINIÇÃO. *Seja S um subconjunto de um espaço euclidiano V . Um elemento de V diz-se ser ortogonal a S se for ortogonal a todo o elemento de S . O conjunto de todos os elementos ortogonais a S representa-se por S^\perp e diz-se “ S perpendicular”.*

É um exercício simples a verificação de que S^\perp é um subespaço de V , quer S o seja ou não. Na hipótese em que S é um subespaço, então S^\perp chama-se o *complemento ortogonal* de S .

EXEMPLO. Se S é um plano passando pela origem, como se indica na figura 1.2, então S^\perp é uma recta passando pela origem e perpendicular a este plano. Este exemplo constitui também uma interpretação geométrica para o teorema que enunciamos a seguir.

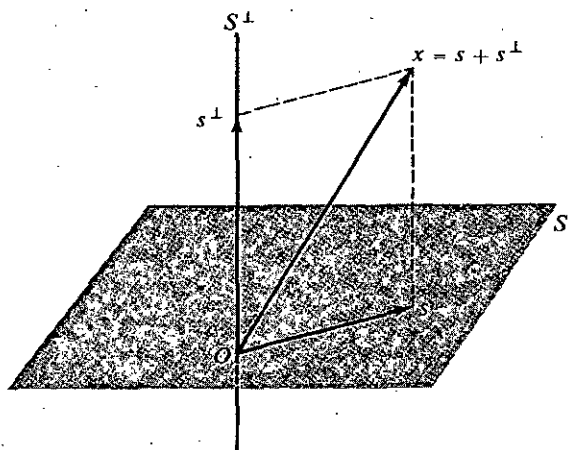


FIG. 1.2. Interpretação geométrica do teorema da decomposição ortogonal em V_3 .

TEOREMA 1.15. TEOREMA DA DECOMPOSIÇÃO ORTOGONAL. *Se V é um espaço euclidiano e S um subespaço de V com dimensão finita, então todo o elemento x de V pode ser representado, de uma única maneira, como a soma de dois elementos, um pertencente a S e outro a S^\perp , isto é, tem-se*

$$(1.17) \quad x = s + s^\perp, \quad \text{onde } s \in S \quad \text{and} \quad s^\perp \in S^\perp.$$

Além disso, a norma de x é dada pela fórmula de Pitágoras

$$(1.18) \quad \|x\|^2 = \|s\|^2 + \|s^\perp\|^2.$$

Demonstração. Em primeiro lugar provamos que é possível a decomposição ortogonal (1.17). Visto que S tem dimensão finita, admite uma base ortonormada, por exemplo $\{e_1, e_2, \dots, e_n\}$. Dado x , definimos os elementos s e s^\perp do modo seguinte:

$$(1.19) \quad s = \sum_{i=1}^n (x, e_i) e_i, \quad s^\perp = x - s.$$

Observa-se que cada termo $(x, e_i) e_i$ é a projecção de x sobre e_i . O elemento s é a soma das projecções de x sobre cada um dos elementos da base. Visto ser s uma combinação linear de elementos da base, s está em S . A definição de s^\perp mostra que (1.17) é verdadeira. Para provar que s^\perp está em S^\perp , consideramos o produto interno de s^\perp com qualquer elemento da base e_j . Temos

$$(s^\perp, e_j) = (x - s, e_j) = (x, e_j) - (s, e_j).$$

Mas de (1.19) resulta que $(s, e_j) = (x, e_j)$, pelo que s^\perp é ortogonal a e_j , o que significa que s^\perp é ortogonal a todo elemento de S , ou seja que $s^\perp \in S^\perp$.

Demonstremos agora que a decomposição (1.17) é única. Admitamos a existência de duas representações para x , por exemplo

$$(1.20) \quad x = s + s^\perp \quad \text{e} \quad x = t + t^\perp,$$

onde s e t pertencem a S e s^\perp e t^\perp pertencem a S^\perp . Desejamos provar que $s = t$ e $s^\perp = t^\perp$. De (1.20), temos $s - t = t^\perp - s^\perp$, pelo que necessitamos unicamente demonstrar que $s - t = 0$. Mas $s - t \in S$ e $t^\perp - s^\perp \in S^\perp$ e assim temos que $s - t$ é simultaneamente ortogonal e igual a $t^\perp - s^\perp$. Porque o elemento zero é o único elemento ortogonal a si próprio, deve ser $s - t = 0$ e portanto a decomposição é única.

Finalmente, provaremos que a norma de x é dada pela fórmula de Pitágoras. Temos

$$\|x\|^2 = (x, x) = (s + s^\perp, s + s^\perp) = (s, s) + (s^\perp, s^\perp),$$

sendo os restantes termos nulos uma vez que s e s^\perp são ortogonais e portanto está demonstrado (1.18).

DEFINIÇÃO. *Seja S um subespaço de dimensão finita de um espaço euclidiano V , e seja $\{e_1, \dots, e_n\}$ uma base ortonormada para S . Se $x \in V$, o elemento s definido por*

$$s = \sum_{i=1}^n (x, e_i) e_i$$

diz-se a projecção de x sobre o subespaço S .

Demonstramos seguidamente que a projecção de x sobre S é a solução do problema de aproximação abordado no início desta Secção.

1.16. A melhor aproximação de elementos de um espaço euclidiano por elementos de um subespaço de dimensão finita

TEOREMA 1.16. TEOREMA DE APROXIMAÇÃO. *Se S é um subespaço de dimensão finita de um espaço euclidiano V e x um elemento qualquer de V , então a projecção de x sobre S está mais próxima de x do que qualquer outro elemento de S , isto é, se s é a projecção de x sobre S , tem-se*

$$\|x - s\| \leq \|x - t\|$$

para todo o $t \in S$, verificando-se o sinal de igualdade se e só se $t = s$.

Demonstração. Considerando o teorema 1.15 podemos escrever $x = s + s^\perp$, com $s \in S$ e $s^\perp \in S^\perp$. Então, para qualquer $t \in S$, tem-se

$$x - t = (x - s) + (s - t).$$

Porque $s - t \in S$ e $x - s = s^\perp \in S^\perp$, esta é uma decomposição ortogonal de $x - t$, pelo que a sua norma é dada pela fórmula de Pitágoras

$$\|x - t\|^2 = \|x - s\|^2 + \|s - t\|^2.$$

Mas $\|s - t\|^2 \geq 0$, pelo que se terá $\|x - t\|^2 \geq \|x - s\|^2$, verificando-se a igualdade se e só se $s = t$, o que completa a demonstração.

EXEMPLO 1. *Aproximação de funções contínuas em $[0, 2\pi]$ por polinómios trigonométricos.* Seja $V = C(0, 2\pi)$, o espaço linear de todas as funções reais contínuas no intervalo $[0, 2\pi]$ e definamos um produto interno pela equação $(f, g) = \int_0^{2\pi} f(x)g(x)dx$. Na Secção 1.12 vimos um conjunto ortogonal de funções trigonométricas $\varphi_0, \varphi_1, \varphi_2, \dots$, onde

$$(1.21) \quad \varphi_0(x) = \frac{1}{\sqrt{2\pi}}, \quad \varphi_{2k-1}(x) = \frac{\cos kx}{\sqrt{\pi}}, \quad \varphi_{2k}(x) = \frac{\sin kx}{\sqrt{\pi}}, \quad \text{para } k \geq 1.$$

Os $2n + 1$ elementos $\varphi_0, \varphi_1, \dots, \varphi_{2n}$ geram um subespaço S de dimensão $2n + 1$. Os elementos de S dizem-se *polinómios trigonométricos*.

Se $f \in C(0, 2\pi)$, seja f_n a projecção de f sobre o subespaço S . Então temos

$$(1.22) \quad f_n = \sum_{k=0}^{2n} (f, \varphi_k) \varphi_k, \quad \text{onde} \quad (f, \varphi_k) = \int_0^{2\pi} f(x) \varphi_k(x) dx.$$

Os números (f, φ_k) chamam-se *coeficientes de Fourier* de f . Usando as fórmulas (1.21), podemos pôr (1.22) na forma

$$(1.23) \quad f_n(x) = \frac{1}{2}a_0 + \sum_{k=1}^n (a_k \cos kx + b_k \sin kx),$$

onde

$$a_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos kx dx, \quad b_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin kx dx$$

para $k = 0, 1, 2, \dots, n$. O teorema de aproximação diz-nos que o polinómio trigonométrico (1.23) aproxima f melhor do que qualquer outro polinómio trigonométrico em S , no sentido de que a norma $\|f - f_n\|$ é a menor possível.

EXEMPLO 2. *Aproximação de funções contínuas em $[-1, 1]$ por polinómios de grau $\leq n$.* Seja $V = C(-1, 1)$, o espaço das funções reais contínuas em $[-1, 1]$, e seja $(f, g) = \int_{-1}^1 f(x)g(x)dx$. Os $n + 1$ polinómios de Legendre normalizados $\varphi_0, \varphi_1, \dots, \varphi_n$, tratados na Secção 1.14, geram um subespaço S de dimensão $n + 1$ formado por todos os polinómios de grau $\leq n$. Se $f \in C(-1, 1)$, represente f_n a projecção de f sobre S . Então tem-se

$$f_n = \sum_{k=0}^n (f, \varphi_k) \varphi_k, \quad \text{onde} \quad (f, \varphi_k) = \int_{-1}^1 f(t) \varphi_k(t) dt.$$

Este é o polinómio de grau $\leq n$ para o qual a norma $\|f - f_n\|$ é mínima. Por exemplo, quando $f(x) = \sin \pi x$, os coeficientes (f, φ_k) são dados por

$$(f, \varphi_k) = \int_{-1}^1 \sin \pi t \varphi_k(t) dt.$$

Em particular, tem-se $(f, \varphi_0) = 0$ e

$$(f, \varphi_1) = \int_{-1}^1 \sqrt{\frac{3}{2}} t \sin \pi t dt = \sqrt{\frac{3}{2}} \frac{2}{\pi}.$$

Por conseguinte o polinómio do 1.º grau $f_1(t)$ que mais se aproxima de $\sin \pi t$ no intervalo fechado $[-1, 1]$ é

$$f_1(t) = \sqrt{\frac{3}{2}} \frac{2}{\pi} \varphi_1(t) = \frac{3}{\pi} t.$$

Porque $(f, \varphi_2) = 0$, este é também a melhor aproximação quadrática.

1.17. Exercícios

- Para cada alínea determinar uma base ortonormada para o subespaço de V_3 gerado pelos vectores indicados.
 (a) $x_1 = (1, 1, 1)$, $x_2 = (1, 0, 1)$, $x_3 = (3, 2, 3)$.
 (b) $x_1 = (1, 1, 1)$, $x_2 = (-1, 1, -1)$, $x_3 = (1, 0, 1)$.
- Para cada alínea determinar uma base ortonormada para o subespaço de V_4 gerado pelos vectores indicados.
 (a) $x_1 = (1, 1, 0, 0)$, $x_2 = (0, 1, 1, 0)$, $x_3 = (0, 0, 1, 1)$, $x_4 = (1, 0, 0, 1)$.
 (b) $x_1 = (1, 1, 0, 1)$, $x_2 = (1, 0, 2, 1)$, $x_3 = (1, 2, -2, 1)$.
- No espaço linear real $C(0, \pi)$, com o produto interno $(x, y) = \int_0^\pi x(t)y(t)dt$, seja $x_n(t) = \cos nt$ para $n = 0, 1, 2, \dots$. Provar que as funções y_0, y_1, y_2, \dots , dadas por

$$y_0(t) = \frac{1}{\sqrt{\pi}} \quad \text{e} \quad y_n(t) = \sqrt{\frac{2}{\pi}} \cos nt \quad \text{para } n \geq 1,$$

formam um conjunto ortonormado gerando o mesmo subespaço que x_0, x_1, x_2, \dots .

- No espaço linear de todos os polinómios reais, com produto interno $(x, y) = \int_0^1 x(t)y(t)dt$, seja $x_n(t) = t^n$ para $n = 0, 1, 2, \dots$. Provar que as funções

$$y_0(t) = 1, \quad y_1(t) = \sqrt{3}(2t - 1), \quad y_2(t) = \sqrt{5}(6t^2 - 6t + 1)$$

formam um conjunto ortonormado gerando o mesmo subespaço que $\{x_0, x_1, x_2\}$.

- Seja V o espaço linear de todas as funções reais f contínuas em $[0, +\infty)$ e tais que o integral $\int_0^\infty e^{-t^2} f(t) dt$ convirja. Definindo $(f, g) = \int_0^\infty e^{-t^2} f(t)g(t) dt$, seja y_0, y_1, y_2, \dots , o conjunto obtido por aplicação do método de Gram-Schmidt a x_0, x_1, x_2, \dots , onde $x_n(t) = t^n$ para $n \geq 0$. Provar que $y_0(t) = 1$, $y_1(t) = t - 1$, $y_2(t) = t^2 - 4t + 2$, $y_3(t) = t^3 - 9t^2 + 18t - 6$.
- No espaço linear real $C(1, 3)$ com produto interno $(f, g) = \int_1^3 f(x)g(x)dx$, seja $f(x) = 1/x$ e demonstrar que o polinómio constante g mais próximo de f é $g = \frac{1}{2} \log 3$. Calcular $\|g - f\|^2$ para este g .
- No espaço linear real $C(0, 2)$ com produto interno $(f, g) = \int_0^2 f(x)g(x)dx$, seja $f(x) = e^x$ e mostrar que o polinómio constante g mais próximo de f é $g = \frac{1}{2}(e^2 - 1)$. Calcular $\|f - g\|^2$ para este g .
- No espaço linear real $C(-1, 1)$ com produto interno $(f, g) = \int_{-1}^1 f(x)g(x)dx$, seja $f(x) = e^x$ e determinar o polinómio do primeiro grau g mais próximo de f . Calcular $\|g - f\|^2$ para este g .
- No espaço linear real $C(0, 2\pi)$ com produto interno $(f, g) = \int_0^{2\pi} f(x)g(x)dx$, seja $f(x) = x$. No subespaço gerado por $u_0(x) = 1$, $u_1(x) = \cos x$, $u_2(x) = \sin x$, determinar o polinómio trigonométrico mais próximo de f .
- No espaço linear V do Exercício 5, seja $f(x) = e^{-x}$ e determinar o polinómio de grau um que mais se aproxima de f .

TRANSFORMAÇÕES LINEARES E MATRIZES

2.1. Transformações lineares

Um dos últimos objectivos da Análise é um amplo estudo de funções cujo domínio e contradomínio são subconjuntos de espaços lineares. Tais funções chamam-se *transformações*, *aplicações* ou *operadores*. Este capítulo trata dos exemplos mais simples, chamados *transformações lineares*, as quais aparecem em todos os ramos da Matemática. As propriedades de transformações mais gerais obtêm-se frequentemente aproximando-as por transformações lineares.

Em primeiro lugar vamos introduzir algumas notações e terminologia relativas a funções quaisquer. Sejam V e W dois conjuntos. O símbolo

$$T: V \rightarrow W$$

será usado para indicar que T é uma função cujo domínio é V e cujos valores estão em W . Para cada x de V , o elemento $T(x)$ em W chama-se a *imagem de x por meio de aplicação T* , e dizemos que T aplica x em $T(x)$. Se A é um subconjunto qualquer de V , o conjunto de todas as imagens $T(x)$, para x em A , chama-se a *imagem de A por meio da aplicação T* e representa-se por $T(A)$. A imagem do domínio V , $T(V)$, é o contradomínio de T .

Suponhamos agora que V e W são espaços lineares admitindo o mesmo conjunto de escalares e definamos uma transformação linear do modo seguinte:

DEFINIÇÃO. Se V e W são espaços lineares, uma função $T: V \rightarrow W$ diz-se uma *transformação linear de V em W* se possui as duas propriedades seguintes:

- (a) $T(x + y) = T(x) + T(y)$ quaisquer que sejam x e y em V ;
- (b) $T(cx) = cT(x)$ para todo o x de V e todo o escalar c .

Estas propriedades significam que T preserva a adição e a multiplicação por escalares. As duas propriedades podem combinar-se numa única fórmula que estabelece que

$$T(ax + by) = aT(x) + bT(y)$$

para todo o par (x, y) em V e quaisquer que sejam os escalares a e b . Por indução, temos também a relação mais geral

$$T\left(\sum_{i=1}^n a_i x_i\right) = \sum_{i=1}^n a_i T(x_i)$$

para n elementos quaisquer x_1, \dots, x_n de V e n escalares quaisquer a_1, \dots, a_n .

O leitor comprovará com facilidade que os exemplos seguintes são transformações lineares.

EXEMPLO 1. *A transformação identidade.* A transformação $T: V \rightarrow V$, onde $T(x) = x$ para todo o x em V , chama-se a transformação identidade e representa-se por I ou por I_v .

EXEMPLO 2. *A transformação zero.* A transformação $T: V \rightarrow V$ que aplica cada elemento de V em O chama-se a transformação zero e representa-se por O .

EXEMPLO 3. *Multiplicação por um escalar fixo c .* Aqui tem-se $T: V \rightarrow V$, onde $T(x) = cx$, para todo o x de V . Quando $c = 1$, cai-se na transformação identidade. Quando $c = 0$, é a transformação zero.

EXEMPLO 4. *Equações lineares.* Sejam $V = V_n$ e $W = V_m$. Dados mn números reais a_{ik} , onde $i = 1, 2, \dots, m$ e $k = 1, 2, \dots, n$, definamos $T: V_n \rightarrow V_m$ do modo seguinte: T aplica cada vector $x = (x_1, \dots, x_n)$ de V_n no vector $y = (y_1, y_2, \dots, y_m)$ de V_m segundo às equações

$$y_i = \sum_{k=1}^n a_{ik} x_k \quad \text{para } i = 1, 2, \dots, m.$$

EXEMPLO 5. *Produto interno com um elemento fixo.* Seja V um espaço euclidiano. Para um elemento fixo z de V , definamos $T: V \rightarrow \mathbf{R}$ do modo seguinte: Se $x \in V$, então $T(x) = (x, z)$, o produto interno de x com z .

EXEMPLO 6. *Projectão sobre um subespaço.* Seja V um espaço euclidiano e S um subespaço de V com dimensão finita. Defina-se $T: V \rightarrow S$ do modo seguinte: Se $x \in V$, então $T(x)$ é a projectão de x sobre S .

EXEMPLO 7. *O operador derivação.* Seja V o espaço linear de todas as funções reais f deriváveis num intervalo aberto (a, b) . A transformação linear que aplica cada função f de V na sua derivada f' chama-se o operador derivação e representa-se por D . Assim temos $D: V \rightarrow W$, onde $D(f) = f'$ para todo o f em V . O espaço W é formado por todas as derivadas f' .

EXEMPLO 8. *O operador integração.* Seja V o espaço linear de todas as funções reais contínuas num intervalo $[a, b]$. Se $f \in V$, defina-se $g = T(f)$ como sendo aquela função de V definida por

$$g(x) = \int_a^x f(t) dt \quad \text{se } a \leq x \leq b.$$

Esta transformação T chama-se o operador integração.

2.2. Espaço nulo e contradomínio

Nesta Secção, T representa uma transformação linear de um espaço linear V em um espaço linear W .

TEOREMA 2.1. *O conjunto $T(V)$ (o contradomínio de T) é um subespaço de W . Além disso, T aplica o elemento zero de V no elemento zero de W .*

Demonstração: Para demonstrar que $T(V)$ é um subespaço de W , necessitamos verificar unicamente os axiomas de fecho. Tomemos dois quaisquer elementos de $T(V)$, por exemplo $T(x)$ e $T(y)$. Então $T(x) + T(y) = T(x + y)$, pelo que $T(x) + T(y)$ está em $T(V)$. Também, para qualquer escalar c temos $cT(x) = T(cx)$, pelo que $cT(x)$ está em $T(V)$. Deste modo, $T(V)$ é um subespaço de W . Fazendo $c = 0$ na relação $T(cx) = cT(x)$, verificamos que $T(O) = O$.

DEFINIÇÃO. *O conjunto de todos os elementos de V que T aplica em O chama-se o espaço nulo de T e representa-se por $N(T)$. Assim tem-se*

$$N(T) = \{x \mid x \in V \quad \text{e} \quad T(x) = O\}.$$

O espaço nulo designa-se também por núcleo de T .

TEOREMA 2.2. *O espaço nulo de T é um subespaço de V .*

Demonstração. Se x e y estão em $N(T)$, o mesmo se verifica com $x + y$ e cx qualquer que seja c , já que

$$T(x + y) = T(x) + T(y) = O \quad \text{e} \quad T(cx) = cT(x) = O.$$

Os exemplos apresentados a seguir referem-se aos espaços nulos das transformações lineares dadas na Secção 2.1.

EXEMPLO 1. *A transformação identidade.* O espaço nulo é $\{O\}$, o subespaço consistindo unicamente do elemento zero.

EXEMPLO 2. *A transformação zero.* Visto cada elemento de V ser aplicado no elemento zero, o espaço nulo é o próprio V .

EXEMPLO 3. *Multiplicação por um escalar fixo c .* Se $c \neq 0$, o espaço nulo contém unicamente O . Se $c = 0$, o espaço nulo é V .

EXEMPLO 4. *Equações lineares.* O espaço nulo consiste de todos os vectores (x_1, \dots, x_n) de V_n para os quais

$$\sum_{k=1}^n a_{ik} x_k = 0 \quad \text{para } i = 1, 2, \dots, m.$$

EXEMPLO 5. *Produto interno com um elemento fixo z .* O espaço nulo consiste de todos os elementos de V ortogonais a z .

EXEMPLO 6. *Projectão sobre um subespaço S .* Se $x \in V$, tem-se a única decomposição ortogonal $x = s + s^\perp$ (pelo teorema (1.15)). Visto ser $T(x) = s$, tem-se $T(x) = O$ se e só se $x = s^\perp$, e assim o espaço nulo é S^\perp , o complemento ortogonal de S .

EXEMPLO 7. *Operador derivação.* O espaço nulo é formado por todas as funções que são constantes num dado intervalo.

EXEMPLO 8. *Operador integração.* O espaço nulo contém unicamente a função zero.

2.3. Nulidade e ordem

Nesta Secção T representa ainda uma transformação linear de um espaço linear V num espaço linear W . Interessa-nos estabelecer uma relação entre a dimensão de V , do espaço nulo $N(T)$ e do contradomínio $T(V)$. Se V tem dimensão finita, então o espaço nulo também tem dimensão finita, porque é um subespaço de V . A dimensão de $N(T)$ chama-se a *nulidade* de T (dimensão do núcleo de T). No teorema que se segue prova-se que a contradomínio também tem dimensão finita; a essa dimensão dá-se o nome de *ordem* de T .

TEOREMA 2.3 **TEOREMA DE NULIDADE MAIS ORDEM.** *Se V é de dimensão finita, então $T(V)$ é também de dimensão finita e tem-se*

$$(2.1) \quad \dim N(T) + \dim T(V) = \dim V.$$

Por outras palavras, a nulidade mais a ordem de uma transformação linear é igual à dimensão do seu domínio.

Demonstração. Sejam $n = \dim V$ e e_1, e_2, \dots, e_k uma base para $N(T)$, onde $k = \dim N(T) \leq n$. Pelo teorema 1.7, estes elementos formam uma parte de uma certa base de V , por exemplo a base

$$(2.2) \quad e_1, \dots, e_k, e_{k+1}, \dots, e_{k+r},$$

com $k + r = n$. Pretendemos provar que os r elementos

$$(2.3) \quad T(e_{k+1}), \dots, T(e_{k+r})$$

formam uma base para $T(V)$, o que prova que $\dim T(V) = r$. Uma vez que $k + r = n$, isto também prova (2.1).

Demonstremos primeiro que os r elementos em (2.3) geram $T(V)$. Se $y \in T(V)$, temos $y = T(x)$ para algum x em V , e podemos escrever $x = c_1 e_1 + \dots + c_{k+r} e_{k+r}$. Daqui resulta

$$y = T(x) = \sum_{i=1}^{k+r} c_i T(e_i) = \sum_{i=1}^k c_i T(e_i) + \sum_{i=k+1}^{k+r} c_i T(e_i) = \sum_{i=k+1}^{k+r} c_i T(e_i)$$

visto que $T(e_1) = \dots = T(e_k) = O$, o que demonstra que os elementos em (2.3) geram $T(V)$.

Vamos agora demonstrar que estes elementos são independentes. Admitamos a existência de escalares c_{k+1}, \dots, c_{k+r} tais que

$$\sum_{i=k+1}^{k+r} c_i T(e_i) = O.$$

A hipótese anterior implica que

$$T\left(\sum_{i=k+1}^{k+r} c_i e_i\right) = O$$

pelo que o elemento $x = c_{k+1} e_{k+1} + \dots + c_{k+r} e_{k+r}$ pertence ao espaço nulo $N(T)$. Quer isto dizer que existem escalares c_1, \dots, c_k , tais que $x = c_1 e_1 + \dots + c_k e_k$, pelo que se terá

$$x - x = \sum_{i=1}^k c_i e_i - \sum_{i=k+1}^{k+r} c_i e_i = O.$$

Mas uma vez que os elementos em (2.2) são independentes, isto implica que todos os escalares c_i são nulos. Consequentemente, os elementos (2.3) são independentes.

Nota: Se V tem dimensão infinita, então pelo menos um dos dois $N(T)$ ou $T(V)$ é de dimensão infinita. Uma demonstração disto é apresentada no Exercício 30 da Secção 2.4.

2.4 Exercícios

Em cada um dos Exercícios 1 a 10, define-se uma transformação $T: V_2 \rightarrow V_2$, mediante a fórmula dada para $T(x, y)$, onde (x, y) é um ponto arbitrário de V_2 . Em cada problema determinar se T é linear e, em caso afirmativo, definir o correspondente espaço nulo e contradomínio, e calcular as suas nulidade e ordem.

1. $T(x, y) = (y, x)$.
2. $T(x, y) = (x, -y)$.
3. $T(x, y) = (x, 0)$.
4. $T(x, y) = (x, x)$.
5. $T(x, y) = (x^2, y^2)$.
6. $T(x, y) = (e^x, e^y)$.
7. $T(x, y) = (x, 1)$.
8. $T(x, y) = (x + 1, y + 1)$.
9. $T(x, y) = (x - y, x + y)$.
10. $T(x, y) = (2x - y, x + y)$.

Fazer o mesmo em cada um dos Exercícios 11 a 15, se a transformação $T: V_2 \rightarrow V_2$ for a que se indica.

11. T imprime a cada ponto uma rotação de um mesmo ângulo ϕ em torno da origem, isto é, T aplica um ponto de coordenadas polares (r, θ) um ponto com coordenadas polares $(r, \theta + \phi)$, com ϕ constante. Além disso T aplica O em si próprio.
12. T aplica cada ponto no seu simétrico relativamente a uma recta dada passando pela origem.
13. T aplica cada ponto no ponto $(1, 1)$.
14. T aplica cada ponto com coordenadas polares (r, θ) no ponto de coordenadas polares $(2r, \theta)$. Além disso, T aplica O em si próprio.
15. T aplica cada ponto com coordenadas polares (r, θ) num ponto com coordenadas polares $(r, 2\theta)$. Além disso, T aplica O em si próprio.

Fazer o mesmo em cada um dos Exercícios 16 a 23 se a transformação $T: V_3 \rightarrow V_3$ se define pela fórmula dada por $T(x, y, z)$, onde (x, y, z) é um ponto arbitrário de V_3 .

16. $T(x, y, z) = (z, y, x)$.
17. $T(x, y, z) = (x, y, 0)$.
18. $T(x, y, z) = (x, 2y, 3z)$.
19. $T(x, y, z) = (x, y, 1)$.
20. $T(x, y, z) = (x + 1, y + 1, z - 1)$.
21. $T(x, y, z) = (x + 1, y + 2, z + 3)$.
22. $T(x, y, z) = (x, y^2, z^3)$.
23. $T(x, y, z) = (x + z, 0, x + y)$.

Em cada um dos Exercícios 24 a 27, a transformação $T: V \rightarrow V$ é a que se indica. Em cada um deles, determinar se T é linear. Se T for linear, dizer quais são o espaço nulo e contradomínio e calcular a nulidade e a ordem da transformação, quando sejam finitas.

24. Seja V o espaço linear de todos os polinómios reais $p(x)$ de grau $\leq n$. Se $p \in V$, $q = T(p)$ significa que $q(x) = p(x + 1)$ para todo o real x .
25. Seja V o espaço linear de todas as funções reais deriváveis no intervalo aberto $(-1, 1)$. Se $f \in V$, $g = T(f)$, significa que $g(x) = xf'(x)$ para todo o x em $(-1, 1)$.
26. Seja V o espaço linear de todas as funções reais contínuas em $[a, b]$. Se $f \in V$, $g = T(f)$ significa que

$$g(x) = \int_a^b f(t) \sin(x - t) dt \quad \text{para } a \leq x \leq b.$$

27. Seja V o espaço de todas as funções reais duas vezes deriváveis num intervalo aberto (a, b) . se $y \in V$, definir $T(y) = y'' + Py' + Qy$, com P e Q constantes dadas.
28. Seja V o espaço linear de todas as sucessões reais convergentes $\{x_n\}$. Definir uma transformação $T: V \rightarrow V$ do modo seguinte: Se $x = \{x_n\}$ é uma sucessão convergente com limite a , seja $T(x) = \{y_n\}$, onde $y_n = a - x_n$ para $n \geq 1$. Provar que T é linear e dizer quais são o espaço nulo e o contradomínio de T .
29. Seja V o espaço linear de todas as funções reais contínuas no intervalo $[-\pi, \pi]$. Seja S o subconjunto de V formado por todas as funções f que satisfazem às três equações

$$\int_{-\pi}^{\pi} f(t) dt = 0, \quad \int_{-\pi}^{\pi} f(t) \cos t dt = 0, \quad \int_{-\pi}^{\pi} f(t) \sin t dt = 0.$$

(a) Provar que S é um subespaço de V .

(b) Provar que S contém as funções $f(x) = \cos nx$ e $f(x) = \sin nx$ para todo $n = 2, 3, \dots$

(c) Provar que S tem dimensão finita.

Seja $T: V \rightarrow V$ uma transformação linear definida do modo seguinte: Se $f \in V$, $g = T(f)$ significa que

$$g(x) = \int_{-\pi}^{\pi} \{1 + \cos(x - t)\} f(t) dt.$$

(d) Provar que $T(V)$, o contradomínio de T , tem dimensão finita e determinar uma base para $T(V)$.

(e) Determinar o espaço nulo de T .

(f) Achar todo o real $c \neq 0$ e todas as funções f não nulas de V tais que $T(f) = cf$ (Observe-se que uma tal função pertence ao contradomínio de T).

30. Seja $T: V \rightarrow W$ uma transformação linear de um espaço linear V em um espaço linear W . Se V possui dimensão infinita, demonstrar que pelo menos um dos $T(V)$ ou $N(T)$ tem dimensão infinita.

[Sugestão: Suponhamos $\dim N(T) = k$, $\dim T(V) = r$ e sejam e_1, \dots, e_k uma base de $N(T)$ e $e_1, \dots, e_k, e_{k+1}, \dots, e_{k+n}$ elementos independentes de V , com $n > r$. Os elementos $T(e_{k+1}), \dots, T(e_{k+n})$ são dependentes uma vez que $n > r$. Utilizar esta conclusão para obter uma tal contradição.]

2.5. Operações algébricas relativas a transformações lineares

Funções cujos valores pertencem ao espaço linear W podem somar-se entre si ou multiplicarem-se por escalares em W de acordo com a seguinte definição.

DEFINIÇÃO. *Sejam $S: V \rightarrow W$ e $T: V \rightarrow W$ duas funções com um domínio comum V e cujos valores estão num espaço linear W . Se c é qualquer escalar em W , definimos a soma $S + T$ e o produto cT pelas igualdades*

$$(2.4) \quad (S + T)(x) = S(x) + T(x), \quad (cT)(x) = cT(x)$$

para todo o x de V .

Interessa-nos particularmente o caso em que V é também um espaço linear com os mesmos escalares que W . Neste caso designamos por $\mathcal{L}(V, W)$ o conjunto de todas as transformações lineares de V em W .

Se S e T são duas transformações lineares de $\mathcal{L}(V, W)$, é uma questão simples a verificação de que $S + T$ e cT são também transformações lineares de $\mathcal{L}(V, W)$. Mais do que isso, ainda, com as operações acabadas de definir, o conjunto $\mathcal{L}(V, W)$ é ele próprio um novo espaço linear. A transformação zero serve de elemento zero deste espaço e a transformação $(-1)T$ é a simétrica de T . É uma questão imediata a verifi-

cação de que os dez axiomas para um espaço linear são satisfeitos. Portanto, temos o seguinte.

TEOREMA 2.4. *O conjunto $\mathcal{L}(V, W)$ de todas as transformações lineares de V em W é um espaço linear com as operações de adição e multiplicação por escalares definidas em (2.4).*

A operação algébrica mais interessante que se efectua com as transformações lineares é a *composição* ou *multiplicação* de transformações. Esta operação não faz qualquer uso da estrutura algébrica de um espaço linear e pode definir-se, com toda a generalidade, de modo seguinte:

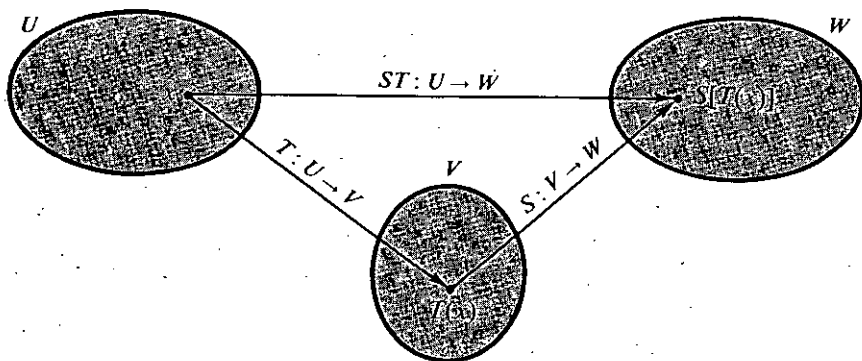


FIG. 16.1. Composição de duas transformações.

DEFINIÇÃO. *Dados os conjuntos U , V e W , sejam $T: U \rightarrow V$ uma função com domínio U e valores em V , e $S: V \rightarrow W$ outra função com domínio V e valores em W . A composição ST é a função $ST: U \rightarrow W$ definida por*

$$(ST)(x) = S[T(x)] \quad \text{para todo } x \text{ em } U$$

Assim, para aplicar x mediante a composição ST , aplicamos em primeiro lugar x por T e depois aplicamos $T(x)$ por S . Isto está representado na figura 2.1.

A composição de funções reais tem-se encontrado repetidas vezes no nosso estudo e vimos que a operação não é, em geral, comutativa. Contudo aqui, como no caso das funções reais, a composição satisfaz a propriedade associativa.

TEOREMA 2.5. *Se $T: U \rightarrow V$, $S: V \rightarrow W$, e $R: W \rightarrow X$ são três funções, tem-se*

$$R(ST) = (RS)T.$$

Demonstração. Ambas as funções $R(ST)$ e $(RS)T$ têm domínio U e valores em X . Para cada x de U , temos

$$[R(ST)](x) = R[(ST)(x)] = R[S[T(x)]] \quad e \quad [(RS)T](x) = (RS)[T(x)] = R[S[T(x)]],$$

o que prova que $R(ST) = (RS)T$.

DEFINIÇÃO. Seja $T: V \rightarrow V$ uma função que aplica V em si próprio. Definem-se as potências inteiras de T por indução do modo seguinte:

$$T^0 = I, \quad T^n = TT^{n-1} \quad \text{para } n \geq 1.$$

Aqui I representa a transformação identidade. O leitor poderá verificar que a propriedade associativa implica a regra $T^m T^n = T^{m+n}$, quaisquer que sejam os inteiros não negativos m e n .

O teorema que se enuncia a seguir mostra que a composição de transformações lineares é ainda linear.

TEOREMA 2.6. Se U, V, W são espaços lineares com os mesmos escalares e se $T: U \rightarrow V$ e $S: V \rightarrow W$ são transformações lineares, então a composição $ST: U \rightarrow W$ é linear.

Demonstração. Para quaisquer x e y de U e quaisquer escalares a e b , temos

$$(ST)(ax + by) = S[T(ax + by)] = S[aT(x) + bT(y)] = aST(x) + bST(y).$$

A composição pode combinar-se com as operações algébricas de adição e multiplicação por escalares em $\mathcal{L}(V, W)$ para dar origem ao seguinte:

TEOREMA 2.7. Sejam U, V e W espaços lineares com os mesmos escalares, suponha-se que S e T pertencem a $\mathcal{L}(V, W)$ e seja c um escalar qualquer.

(a) Para qualquer função R com valores em V , tem-se

$$(S + T)R = SR + TR \quad e \quad (cS)R = c(SR).$$

(b) Para qualquer transformação linear $R: W \rightarrow U$, tem-se

$$R(S + T) = RS + RT \quad e \quad R(cS) = c(RS).$$

A demonstração é uma aplicação imediata da definição de composição e é deixada ao leitor como exercício.

2.6. Inversas

No nosso estudo das funções de uma variável real aprendemos a construir novas funções por inversão das funções monótonas. Pretendemos agora generalizar o processo de inversão a uma classe mais geral de funções.

Dada uma função T , é nosso objectivo encontrar, se possível, outra função S cuja composição com T seja a transformação idêntica. Visto a composição não ser, em geral, comutativa, temos que distinguir TS de ST . Para tal introduzimos duas espécies de funções inversas que chamamos inversa esquerda e inversa direita.

DEFINIÇÃO. Dados dois conjuntos V e W e uma função $T: V \rightarrow W$, diz-se que uma função $S: T(V) \rightarrow V$ é a inversa esquerda de T se $S[T(x)] = x$ para todo o x em V , isto é, se

$$ST = I_V,$$

onde I_V é a transformação identidade em V . Uma função $R: T(V) \rightarrow V$ diz-se inversa direita de T se $T[R(y)] = y$ para todo o y em $T(V)$, isto é, se

$$TR = I_{T(V)},$$

com $I_{T(V)}$ a transformação identidade em $T(V)$.

EXEMPLO. Uma função sem inversa esquerda, mas com duas inversas direitas. Seja $V = \{1, 2\}$ e $W = \{0\}$. Define-se $T: V \rightarrow W$ do modo seguinte: $T(1) = T(2) = 0$. Esta função admite duas inversas direitas $R: W \rightarrow V$ e $R': W \rightarrow V$ definidas por

$$R(0) = 1, \quad R'(0) = 2.$$

Não é possível definir-se uma inversa esquerda S , porque requereria

$$1 = S[T(1)] = S(0) \quad \text{e} \quad 2 = S[T(2)] = S(0).$$

Este exemplo simples põe em evidência que as inversas esquerdas não existem necessariamente e que as inversas direitas não são necessariamente únicas.

Toda a função $T: V \rightarrow W$ tem, pelo menos, uma inversa direita. Com efeito, cada y de $T(V)$ tem a forma $y = T(x)$ para, pelo menos, um x em V . Se escolhemos um tal x e definimos $R(y) = x$, então $T[R(y)] = T(x) = y$ para cada y em $T(V)$, pelo que R é uma inversa direita. A não unicidade pode ocorrer devido a que pode haver mais do que um x de V que se aplique num y de $T(V)$. No teorema 16.9 provaremos que se cada y em $T(V)$ é a imagem de um só x de V , então a inversa direita é única.

Antes porém vamos demonstrar que se existe uma inversa esquerda, ela é única e, ao mesmo tempo, é inversa direita.

TEOREMA 2.8. Uma função $T: V \rightarrow W$ pode ter, quando muito, uma inversa esquerda. Se T tem uma inversa esquerda S , então S é também inversa direita.

Demonstração. Admitamos que T possui duas inversas esquerdas, $S: T(V) \rightarrow V$ e $S': T(V) \rightarrow V$. Escolhamos qualquer y de $T(V)$. Vamos provar que $S(y) = S'(y)$. Como $y = T(x)$ para algum x de V , temos

$$S[T(x)] = x \quad \text{e} \quad S'[T(x)] = x,$$

visto que S e S' são ambas inversas esquerdas. Deste modo $S(y) = x$ e $S'(y) = x$, pelo que $S(y) = S'(y)$ para todo o y em $T(V)$. Por conseguinte $S = S'$, o que prova que a inversa esquerda é única.

Vamos agora demonstrar que toda a inversa esquerda S é também inversa direita. Escolhamos qualquer elemento y em $T(V)$. Pretendemos demonstrar que $T[S(y)] = y$. Uma vez que $y \in T(V)$, temos $y = T(x)$ para algum x em V . Mas S é uma inversa esquerda, pelo que

$$x = S[T(x)] = S(y).$$

Aplicando T , obtemos $T(x) = T[S(y)]$. Mas $y = T(x)$, pelo que $y = T[S(y)]$, o que completa a demonstração.

O teorema que apresentamos a seguir caracteriza todas as funções que admitem inversa esquerda.

TEOREMA 2.9. *Uma função $T: V \rightarrow W$ tem uma inversa esquerda se e só se T aplica elementos distintos de V em elementos distintos de W , isto é, se e só se para quaisquer x e y de V ,*

$$(2.5) \quad x \neq y \quad \text{implica} \quad T(x) \neq T(y).$$

Nota: A condição (2.5) é equivalente à afirmação

$$(2.6) \quad T(x) = T(y) \quad \text{implica} \quad x = y.$$

Uma função T satisfazendo a (2.5) ou (2.6) para quaisquer x e y em V diz-se *biunívoca* em V .

Demonstração. Suponhamos que T tem uma inversa esquerda S e admitamos que $T(x) = T(y)$. Desejamos provar que $x = y$. Aplicando S , encontramos $S[T(x)] = S[T(y)]$. Visto que $S[T(x)] = x$ e $S[T(y)] = y$, isto implica $x = y$, o que prova que uma função com inversa esquerda é biunívoca no seu domínio.

Demonstremos agora a propriedade inversa. Admitamos que T é biunívoca no seu domínio. Pretendemos provar a existência de uma função $S: T(V) \rightarrow V$ que será a inversa esquerda de T . Se $y \in T(V)$, então $y = T(x)$ para algum x em V . Por (2.6), existe precisamente um x em V para o qual $y = T(x)$. Definamos $S(y)$ como sendo esse x . Isto é definamos S sobre $T(V)$ do modo seguinte:

$S(y) = x$ significa que $T(x) = y$.

Então temos $S[T(x)] = x$ para todo o x em V , pelo que $ST = I_V$. Desta maneira, a função S assim definida é a inversa esquerda de T .

DEFINIÇÃO. *Seja $T: V \rightarrow W$ biunívoca em V . A única inversa esquerda de T (que se sabe já ser também inversa direita) representa-se por T^{-1} . Diz-se que T é invertível e chama-se T^{-1} a inversa de T .*

Os resultados desta Secção dizem respeito a funções quaisquer. Vamos em seguida aplicá-los a transformações lineares.

2.7. Transformações lineares biunívocas

Nesta Secção, V e W representam espaços lineares com os mesmos escalares, e $T: V \rightarrow W$ representa uma transformação linear em $\mathcal{L}(V, W)$. A linearidade de T permite-nos exprimir de diversas maneiras a propriedade para que uma transformação linear seja biunívoca.

TEOREMA 2.10. *Seja $T: V \rightarrow W$ uma transformação linear em $\mathcal{L}(V, W)$. São equivalentes as seguintes proposições:*

- (a) T é biunívoca em V .
- (b) T é invertível e a sua inversa $T^{-1}: T(V) \rightarrow V$ é linear.
- (c) Para todo o x em V , $T(x) = O$, implica $x = O$, isto é, o espaço nulo $N(T)$ contém unicamente o elemento zero de V .

Demonstração. Vamos demonstrar que (a) implica (b), (b) implica (c) e (c) implica (a). Admitamos que (a) é verdadeira. Então T admite inversa T^{-1} pelo teorema 2.9, a qual se vai provar que é linear. Consideremos dois quaisquer elementos u e v em $T(V)$. Então $u = T(x)$ e $v = T(y)$ para algum x e algum y em V . Quaisquer que sejam os escalares a e b , temos

$$au + bv = aT(x) + bT(y) = T(ax + by),$$

visto T ser linear. Daqui resulta, aplicando T^{-1} ,

$$T^{-1}(au + bv) = ax + by = aT^{-1}(u) + bT^{-1}(v),$$

pelo que T^{-1} é linear. Por conseguinte (a) implica (b).

Admitamos agora que (b) é verdadeira. Tomemos um x qualquer em V para o qual $T(x) = O$. Aplicando T^{-1} , encontramos que $x = T^{-1}(O) = O$, visto T^{-1} ser linear. Assim concluímos que (b) implica (c).

Finalmente, admitamos que (c) é verdadeira. Consideremos dois quaisquer elementos u e v em V com $T(u) = T(v)$. Devido à linearidade, temos $T(u - v) = T(u) - T(v) =$

O , pelo que $u - v = O$. Portanto, T é biunívoca em V , e a demonstração do teorema está completada.

Quando V tem dimensão finita, a propriedade da transformação ser biunívoca pode ser formulada em termos de independência e dimensionalidade, como se indica no teorema que apresentamos a seguir.

TEOREMA 2.11. *Se $T: V \rightarrow W$ é uma transformação linear em $\mathcal{L}(V, W)$ e V tem dimensão finita, $\dim V = n$, são equivalentes as seguintes proposições:*

- (a) T é biunívoca em V .
- (b) Se e_1, \dots, e_p são elementos independentes de V , então $T(e_1), \dots, T(e_p)$ são elementos independentes de $T(V)$.
- (c) $\dim T(V) = n$.
- (d) Se $\{e_1, \dots, e_n\}$ é uma base de V , então $\{T(e_1), \dots, T(e_n)\}$ é uma base de $T(V)$.

Demonstração. Vamos ainda aqui provar que (a) implica (b), (b) implica (c), (c) implica (d) e (d) implica (a). Admitamos que (a) é verdadeira. Sejam e_1, \dots, e_p elementos independentes de V e consideremos os elementos $T(e_1), \dots, T(e_p)$ em $T(V)$. Admitamos que

$$\sum_{i=1}^p c_i T(e_i) = O$$

para certos escalares c_1, \dots, c_p . Devido à linearidade, obtemos:

$$T\left(\sum_{i=1}^p c_i e_i\right) = O, \quad \text{e por isso} \quad \sum_{i=1}^p c_i e_i = O$$

visto ser T biunívoca. Mas e_1, \dots, e_p são independentes, pelo que $c_1 = \dots = c_p = 0$ e, por conseguinte, (a) implica (b).

Admitamos agora que (b) é verdadeira. Seja $\{e_1, \dots, e_n\}$ uma base de V . Devido a (b), os n elementos $T(e_1), \dots, T(e_n)$ em $T(V)$ são independentes. Deste modo $\dim T(V) \geq n$. Mas, pelo teorema 2.3, temos que $\dim T(V) \leq n$ e, por conseguinte, $\dim T(V) = n$, pelo que (b) implica (c).

Admitamos agora que (c) é válida e seja $\{e_1, \dots, e_n\}$ uma base de V . Consideremos um elemento qualquer y em $T(V)$. Então $y = T(x)$ para algum x em V , pelo que se terá

$$x = \sum_{i=1}^n c_i e_i, \quad \text{e por isso} \quad y = T(x) = \sum_{i=1}^n c_i T(e_i).$$

Portanto $\{T(e_1), \dots, T(e_n)\}$ gera $T(V)$. Mas tínhamos admitido que $\dim T(V) = n$, pelo que $\{T(e_1), \dots, T(e_n)\}$ é uma base de $T(V)$ e, por conseguinte, (c) implica (d).

Finalmente, admitamos que (d) é verdadeira. Provaremos que $T(x) = O$ implica $x = O$. Seja $\{e_1, \dots, e_n\}$ uma base de V . Se $x \in V$, podemos escrever

$$x = \sum_{i=1}^n c_i e_i, \quad \text{e por isso} \quad T(x) = \sum_{i=1}^n c_i T(e_i).$$

Se $T(x) = O$, então $c_1 = \dots = c_n = 0$, visto que os elementos $T(e_1), \dots, T(e_n)$ são independentes. Assim será $x = O$, pelo que T é biunívoca em V , o que significa que (d) implica (a) e o teorema está demonstrado.

2.8. Exercícios

1. Seja $V = \{0, 1\}$. Descrever todas as funções $T: V \rightarrow V$. No total são quatro. Designando-as por T_1, T_2, T_3 e T_4 construir uma tábua de multiplicação que mostre a composição de cada par. Indicar quais das funções são biunívocas em V e definir as correspondentes inversas.
2. Seja $V = \{0, 1, 2\}$. Descrever todas as funções $T: V \rightarrow V$ para as quais $T(V) = V$. Existem seis ao todo. Designando-as por T_1, \dots, T_6 construir uma tábua de multiplicação mostrando a composição de cada par. Dizer quais das funções são biunívocas em V , e definir as suas inversas.

Em cada um dos Exercícios 3 a 12, define-se uma função $T: V_2 \rightarrow V_2$ pela fórmula dada para $T(x, y)$, onde (x, y) é um ponto arbitrário de V_2 . Determinar para cada um se T é biunívoca em V_2 . Caso afirmativo, definir o seu contradomínio $T(V_2)$; para cada ponto (u, v) em $T(V_2)$, seja $(x, y) = T^{-1}(u, v)$ e dar fórmulas que determinem x e y em função de u e v .

- | | |
|-----------------------------|-----------------------------------|
| 3. $T(x, y) = (y, x)$. | 8. $T(x, y) = (e^x, e^y)$. |
| 4. $T(x, y) = (x, -y)$. | 9. $T(x, y) = (x, 1)$. |
| 5. $T(x, y) = (x, 0)$. | 10. $T(x, y) = (x + 1, y + 1)$. |
| 6. $T(x, y) = (x, x)$. | 11. $T(x, y) = (x - y, x + y)$. |
| 7. $T(x, y) = (x^2, y^2)$. | 12. $T(x, y) = (2x - y, x + y)$. |

Em cada um dos Exercícios 13 a 20, define-se uma função $T: V_3 \rightarrow V_3$ pela fórmula dada para $T(x, y, z)$, onde (x, y, z) é um ponto arbitrário de V_3 . Em cada hipótese, determinar se T é biunívoca em V_3 . Caso afirmativo, deduzir o seu contradomínio $T(V_3)$; para cada ponto (u, v, w) em $T(V_3)$, seja $(x, y, z) = T^{-1}(u, v, w)$ e dar fórmulas para a determinação de x, y e z em função de u, v e w .

- | | |
|--|--|
| 13. $T(x, y, z) = (z, y, x)$ | 17. $T(x, y, z) = (x + 1, y + 1, z - 1)$. |
| 14. $T(x, y, z) = (x, y, 0)$. | 18. $T(x, y, z) = (x + 1, y + 2, z + 3)$. |
| 15. $T(x, y, z) = (x, 2y, 3z)$. | 19. $T(x, y, z) = (x, x + y, x + y + z)$. |
| 16. $T(x, y, z) = (x, y, x + y + z)$. | 20. $T(x, y, z) = (x + y, y + z, x + z)$. |

21. Seja $T: V \rightarrow V$ uma função que aplica V sobre si próprio. Inductivamente, definem-se potências pelas fórmulas $T^0 = 1$, $T^n = TT^{n-1}$ para $n \geq 1$. Provar que a propriedade associativa para a composição implica a regra $T^m T^n = T^{m+n}$. Se T é invertível, provar que T^n é também invertível e que $(T^n)^{-1} = (T^{-1})^n$.

Nos Exercícios 22 a 25, S e T representam funções com domínio V e valores em V . Em geral, $ST \neq TS$. Se $ST = TS$, dizemos que S e T comutam (ou permutam).

22. Se S e T comutam, provar que $(ST)^n = S^n T^n$ para quaisquer inteiros $n \geq 0$.
23. Se S e T são invertíveis e comutam, provar que ST é também invertível e que $(ST)^{-1} = T^{-1} S^{-1}$ ou, por outras palavras, a inversa de ST é a composição das inversas, tomadas por ordem inversa.
24. Se S e T são invertíveis e comutam, provar que as respectivas inversas também comutam.

25. Seja V um espaço linear. Se S e T comutam, provar que

$$(S + T)^2 = S^2 + 2ST + T^2 \quad \text{e} \quad (S + T)^3 = S^3 + 3S^2T + 3ST^2 + T^3.$$

Indicar como devem modificar-se estas fórmulas se $ST \neq TS$.

26. Sejam S e T transformações lineares de V_1 em V_1 definidas pelas fórmulas $S(x, y, z) = (z, y, x)$ e $T(x, y, z) = (x, x + y, x + y + z)$, com (x, y, z) um ponto arbitrário de V_1 .
- (a) Determinar a imagem de (x, y, z) sob cada uma das transformações seguintes: ST , TS , $ST - TS$, S^2 , T^2 , $(ST)^2$, $(TS)^2$, $(ST - TS)^2$.
- (b) Provar que S e T são biunívocas em V_1 e determinar a imagem de (u, v, w) sob cada uma das seguintes transformações: S^{-1} , T^{-1} , $(ST)^{-1}$, $(TS)^{-1}$.
- (c) Determinar a imagem de (x, y, z) sob $(T - I)^n$ para cada $n \geq 1$.
27. Seja V o espaço linear de todos os polinómios reais $p(x)$. Represente D o operador derivação e seja T o operador integração que aplica cada polinómio p num polinómio q dado por $q(x) = \int p(t)dt$. Provar que $DT = I_V$ mas que $TD \neq I_V$. Descrever o espaço nulo e o contradomínio de TD .
28. Seja V o espaço linear de todos os polinómios reais $p(x)$. Represente D o operador derivação e seja T a transformação linear que aplica $p(x)$ em $xp(x)$.
- (a) Seja $p(x) = 2 + 3x - x^2 + 4x^3$ e determinamos a imagem de p sob cada uma das seguintes transformações:
 D , T , DT , TD , $DT - TD$, $T^2D^2 - D^2T^2$.
- (b) Determinar os polinómios p de V para os quais $T(p) = p$.
- (c) Determinar os polinómios p de V para os quais $(DT - 2D)(p) = 0$.
- (d) Determinar os polinómios p de V para os quais $(DT - TD)(p) = D^n(p)$.
29. Sejam V e D como no Exercício 28, mas seja T uma transformação linear que aplica $p(x)$ em $x^p(x)$. Provar que $DT - TD = I$ e que $DT^n - T^nD = nT^{n-1}$ para $n \geq 2$.
30. Sejam S e T de $\mathcal{L}(V, V)$ e admita-se que $ST - TS = I$. Provar que $ST^n - T^nS = nT^{n-1}$ para todo $n \geq 1$.
31. Seja V o espaço linear de todos os polinómios reais $p(x)$. Sejam R , S e T funções que aplicam um polinómio arbitrário $p(x) = c_0 + c_1x + \dots + c_nx^n$ de V no polinómio $r(x)$, $s(x)$ e $t(x)$, respectivamente, onde

$$r(x) = p(0), \quad s(x) = \sum_{k=1}^n c_k x^{k-1}, \quad t(x) = \sum_{k=0}^n c_k x^{k+1}.$$

- (a) Seja $p(x) = 2 + 3x - x^2 + x^3$ e determinar a imagem de p sob cada uma das seguintes transformações: R , S , T , ST , TS , $(TS)^2$, T^2S^2 , S^2T^2 , TRS , RST .
- (b) Provar que R , S e T são lineares e determinar o espaço nulo e o contradomínio de cada uma.
- (c) Provar que T é biunívoca em V e determinar a sua inversa.
- (d) Se $n \geq 1$, exprimir $(TS)^n$ e S^nT^n em função de I e R .
32. Considerar o Exercício 28 da Secção 2.4. Determinar se T é biunívoca em V . Caso afirmativo, dizer qual é a sua inversa.

2.9. Transformações lineares com valores determinados

Se V tem dimensão finita, podemos sempre construir uma transformação linear $T: V \rightarrow W$ com valores determinados para os elementos duma base de V , como se explica no seguinte:

TEOREMA 2.12. *Se e_1, \dots, e_n constitui uma base de um espaço linear n -dimensional V e u_1, \dots, u_n são n elementos arbitrários de um espaço linear W , então existe uma e uma só transformação linear $T: V \rightarrow W$ tal que*

$$(2.7) \quad T(e_k) = u_k \quad \text{para } k = 1, 2, \dots, n.$$

Esta transformação T aplica um elemento arbitrário x de V do modo seguinte:

$$(2.8) \quad \text{Se } x = \sum_{k=1}^n x_k e_k, \quad \text{então } T(x) = \sum_{k=1}^n x_k u_k.$$

Demonstração. Cada x de V pode exprimir-se de uma única maneira como combinação linear de e_1, \dots, e_n , sendo os coeficientes x_1, \dots, x_n as componentes de x na base ordenada (e_1, \dots, e_n) . Se definimos T por (2.8), é uma questão imediata a verificação de que T é linear. Se $x = e_k$ para certo k , então todas as componentes de x são 0, excepto a de ordem k , que é 1, pelo que (2.8) dá $T(e_k) = u_k$, como se pretendia provar.

Para demonstrar que existe unicamente uma transformação linear satisfazendo (2.7), designamos por T' outra transformação e calculamos $T'(x)$. Encontramos

$$T'(x) = T' \left(\sum_{k=1}^n x_k e_k \right) = \sum_{k=1}^n x_k T'(e_k) = \sum_{k=1}^n x_k u_k = T(x).$$

Visto que $T'(x) = T(x)$ para todo o x em V , temos $T' = T$, o que completa a demonstração.

EXEMPLO. Determinar a transformação linear $T: V_2 \rightarrow V_2$ que aplica os elementos base $i = (1, 0)$ e $j = (0, 1)$ do modo seguinte:

$$T(i) = i + j, \quad T(j) = 2i - j.$$

Resolução. Se $x = x_1 i + x_2 j$ é um elemento arbitrário de V_2 , então $T(x)$ é dada por

$$T(x) = x_1 T(i) + x_2 T(j) = x_1(i + j) + x_2(2i - j) = (x_1 + 2x_2)i + (x_1 - x_2)j.$$

2.10. Representação matricial das transformações lineares

O teorema 2.12 mostra que uma transformação linear $T: V \rightarrow W$ de um espaço linear de dimensão finita V fica completamente determinada pela sua acção sobre um dado conjunto de elementos de uma base e_1, e_2, \dots, e_n de V . Suponhamos agora que o espaço W tem também dimensão finita, por exemplo $\dim W = m$ e seja w_1, \dots, w_m uma base de W . (As dimensões m e n podem não ser iguais.) Visto T ter valores em W , cada elemento $T(e_k)$ pode representar-se de maneira única, como uma combinação linear dos elementos de base w_1, w_2, \dots, w_m , a saber

$$T(e_k) = \sum_{i=1}^m t_{ik} w_i,$$

onde t_{1k}, \dots, t_{mk} são as componentes de $T(e_k)$ na base ordenada (w_1, w_2, \dots, w_m) . Dispostemos verticalmente o m -sistema (t_{1k}, \dots, t_{mk}) do modo seguinte:

$$(2.9) \quad \begin{bmatrix} t_{1k} \\ t_{2k} \\ \vdots \\ t_{mk} \end{bmatrix}.$$

Esta disposição chama-se *um vector coluna* ou *uma matriz coluna*. Teremos um tal vector coluna para cada um dos n elementos $T(e_1), \dots, T(e_n)$. Colocamo-los lado a lado, encerrando-os por um par de parêntesis rectos de modo a obter-se a seguinte disposição rectangular.

$$\begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ t_{21} & t_{22} & \cdots & t_{2n} \\ \vdots & \vdots & & \vdots \\ t_{m1} & t_{m2} & \cdots & t_{mn} \end{bmatrix}.$$

Este arranjo diz-se uma *matriz* formada por m linhas e n colunas. Chamamo-la uma matriz m por n ou uma matriz $m \times n$. A primeira linha é uma matriz $1 \times n$ ($t_{11}, t_{12}, \dots, t_{1n}$). A matriz $m \times 1$ destacada em (2.9) é a coluna de ordem k . Os escalares t_{ik} estão afectados de dois índices, indicando o primeiro (o índice i) a *linha*, e o segundo (o índice k) a *coluna* em que se situa t_{ik} . Chamamos a t_{ik} o *elemento ik* da matriz. Também se utiliza por vezes uma notação mais compacta

$$(t_{ik}), \quad \text{ou} \quad (t_{ik})_{i,k=1}^{m,n},$$

para apresentar a matriz cujo elemento ik é t_{ik} .

Assim, cada transformação linear T de um espaço n dimensional V sobre um espaço m dimensional W dá lugar a uma matriz $m \times n$, (t_{ik}) , cujas colunas são as componentes de $T(e_1), \dots, T(e_n)$ relativamente à base (w_1, \dots, w_m) . A matriz considerada define a *representação matricial* de T relativamente à escolha de bases ordenadas (e_1, \dots, e_n) para V e (w_1, \dots, w_m) para W . Uma vez conhecida a matriz (t_{ik}) , as componentes de qualquer elemento $T(x)$ relativamente à base (w_1, \dots, w_m) pode determinar-se como se indica a seguir.

TEOREMA 2.13. *Seja T uma transformação linear em $\mathcal{L}(V, W)$, onde $\dim V = n$ e $\dim W = m$ e sejam (e_1, \dots, e_n) e (w_1, \dots, w_m) bases ordenadas de V e W , respectivamente, e seja (t_{ik}) uma matriz $m \times n$ cujos elementos são definidos pelas equações*

$$(2.10) \quad T(e_k) = \sum_{i=1}^m t_{ik} w_i, \quad \text{para } k = 1, 2, \dots, n.$$

Então um elemento arbitrário

$$(2.11) \quad x = \sum_{k=1}^n x_k e_k$$

de V com componentes (x_1, \dots, x_n) relativamente a (e_1, \dots, e_n) é aplicado por T num elemento

$$(2.12) \quad T(x) = \sum_{i=1}^m y_i w_i$$

de W com componentes (y_1, \dots, y_m) relativamente a (w_1, \dots, w_m) . Os y_i estão relacionados com as componentes de x mediante as equações lineares

$$(2.13) \quad y_i = \sum_{k=1}^n t_{ik} x_k \quad \text{para } i = 1, 2, \dots, m.$$

Demonstração. Aplicando T a cada membro de (2.11) e considerando (2.10), obtemos

$$T(x) = \sum_{k=1}^n x_k T(e_k) = \sum_{k=1}^n x_k \sum_{i=1}^m t_{ik} w_i = \sum_{i=1}^m \left(\sum_{k=1}^n t_{ik} x_k \right) w_i = \sum_{i=1}^m y_i w_i,$$

onde cada y_i é dado por (2.13), o que completa a demonstração.

Tendo escolhido um par de bases (e_1, \dots, e_n) e (w_1, \dots, w_m) para V e W , respectivamente, toda a transformação linear $T: V \rightarrow W$ admite numa representação matricial (t_{ik}) . Inversamente, se dispomos mn escalares conforme os elementos da matriz (t_{ik}) e

escolhemos um par de bases ordenadas para V e W , então é fácil provar que existe precisamente uma transformação linear $T: V \rightarrow W$ tendo aquela representação matricial. Definimos muito simplesmente T com os elementos base de V por intermédio de (2.10). Então, pelo teorema 2.12, existe uma e uma só transformação linear $T: V \rightarrow W$ com aqueles valores previamente determinados. A imagem $T(x)$ de um ponto arbitrário x de V é então dada pelas equações (2.12) e (2.13).

EXEMPLO 1. *Construção de uma dada transformação linear a partir de uma matriz dada.* Suponhamos que partimos com a matriz 2×3 de elementos

$$\begin{bmatrix} 3 & 1 & -2 \\ 1 & 0 & 4 \end{bmatrix}.$$

Escolhamos as bases usuais de vectores unitários coordenados para V_3 e V_2 . Então, a matriz dada representa uma transformação linear $T: V_3 \rightarrow V_2$ a qual aplica um vector arbitrário (x_1, x_2, x_3) de V_3 no vector (y_1, y_2) de V_2 segundo as equações lineares

$$y_1 = 3x_1 + x_2 - 2x_3,$$

$$y_2 = x_1 + 0x_2 + 4x_3.$$

EXEMPLO 2. *Construção da representação matricial de uma transformação linear dada.* Seja V o espaço linear de todos os polinómios reais $p(x)$ de grau ≤ 3 . Este espaço tem dimensão 4, e escolhemos a base $(1, x, x^2, x^3)$. Seja D o operador derivação que aplica cada polinómios $p(x)$ em V na sua derivada $p'(x)$. Podemos considerar D como uma transformação linear de V em W , onde W é o espaço tridimensional de todos os polinómios reais de grau ≤ 2 . Em W escolhemos a base $(1, x, x^2)$. Para determinar a representação matricial de D , relativamente a esta escolha de bases, transformamos (derivamos) cada elemento da base de V e exprimimo-lo como uma combinação linear dos elementos da base de W . Assim, encontramos

$$D(1) = 0 = 0 + 0x + 0x^2, \quad D(x) = 1 = 1 + 0x + 0x^2,$$

$$D(x^2) = 2x = 0 + 2x + 0x^2, \quad D(x^3) = 3x^2 = 0 + 0x + 3x^2.$$

Os coeficientes destes polinómios determinam as *colunas* da representação matricial de D . Deste modo, a representação pedida vem dada pela matriz 3×4 :

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}.$$

Para evidenciar o facto de que a representação matricial depende não só dos elementos das bases mas também da respectiva ordem, invertamos a ordem dos elementos da base de W e utilizemos, em seu lugar, a base ordenada $(x^2, x, 1)$. Então os elementos da base de V são transformados nos mesmos polinómios obtidos atrás, mas as componentes destes polinómios relativamente à nova base $(x^2, x, 1)$ aparecem por ordem inversa. Portanto, a representação matricial de D vem agora

$$\begin{bmatrix} 0 & 0 & 0 & 3 \\ 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

Calculemos uma terceira representação matricial para D , usando a base $(1, 1+x, 1+x+x^2, 1+x+x^2+x^3)$ para V , e a base $(1, x, x^2)$ para W . Os elementos da base de V transformam-se do modo seguinte:

$$D(1) = 0, \quad D(1+x) = 1, \quad D(1+x+x^2) = 1+2x,$$

$$D(1+x+x^2+x^3) = 1+2x+3x^2,$$

pelo que a representação matricial vem, neste caso,

$$\begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 \end{bmatrix}.$$

2.11. Construção de uma representação matricial na forma diagonal

Uma vez que é possível obter diferentes representações matriciais de uma dada transformação linear por diferentes escolhas de bases, é natural tentar a escolha de bases de maneira que a matriz resultante tenha uma forma particularmente simples. O teorema que enunciaremos a seguir mostra que podemos fazer com que todos os elementos da matriz sejam nulos, excepto possivelmente ao longo da diagonal partindo do canto superior esquerdo da matriz. Esta diagonal será formada por uma fileira de elementos 1 seguidos de zeros, sendo o número de 1 igual à ordem da transformação. Uma matriz (t_{ik}) com todos os elementos $t_{ik} = 0$ quando $i \neq k$ chama-se *matriz diagonal*.

TEOREMA 2.14. *Sejam V e W espaços lineares de dimensão finita, com $\dim V = n$ e $\dim W = m$. Admita-se que $T \in \mathcal{L}(V, W)$ e represente $r = \dim T(V)$ a ordem de T . Existem então uma base para V (e_1, \dots, e_n) , e uma base (w_1, \dots, w_n) para W tais que*

$$(2.14) \quad T(e_i) = w_i \quad \text{para} \quad i = 1, 2, \dots, r,$$

e

$$(2.15) \quad T(e_i) = 0 \quad \text{para} \quad i = r+1, \dots, n.$$

Por conseguinte, a matriz (t_{ik}) de T relativa a estas bases tem todos os elementos zero, excepto para os r elementos da diagonal

$$t_{11} = t_{22} = \dots = t_{rr} = 1.$$

Demonstração: Construimos, em primeiro lugar, uma base para W . Porque $T(V)$ é um subespaço de W com $\dim T(V) = r$, o espaço $T(V)$ tem uma base de r elementos em W , sejam w_1, w_2, \dots, w_r . Pelo teorema 1.7, estes elementos formam um subconjunto de uma certa base de W . Deste modo podemos juntar os elementos w_{r+1}, \dots, w_m de modo que

$$(2.16) \quad (w_1, \dots, w_r, w_{r+1}, \dots, w_m)$$

seja uma base de W .

Seguidamente construímos uma base para V . Cada um dos primeiros r elementos w_i em (2.16) é a imagem de pelo menos um elemento de V . Escolhamos um tal elemento de V e designemo-lo por e_i . Então $T(e_i) = w_i$ para $i = 1, 2, \dots, r$ pelo que (2.14) é satisfeita. Seja agora k a dimensão do espaço nulo $N(T)$. Pelo teorema 2.3 temos $n = k + r$. Visto ser $\dim N(T) = k$, o espaço $N(T)$ admite uma base formada por k elementos de V , que designamos por e_{r+1}, \dots, e_{r+k} . Para cada um destes elementos, a equação (16.15) é satisfeita. Portanto, para completar a demonstração, devemos provar que o conjunto ordenado

$$(2.17) \quad (e_1, \dots, e_r, e_{r+1}, \dots, e_{r+k})$$

é uma base para V . Porque $\dim V = n = r + k$, necessitamos unicamente mostrar que estes elementos são independentes. Suponhamos que certa combinação linear deles seja zero, por exemplo

$$(2.18) \quad \sum_{i=1}^{r+k} c_i e_i = 0.$$

Aplicando T e usando as equações (2.14) e (2.15) encontramos

$$\sum_{i=1}^{r+k} c_i T(e_i) = \sum_{i=1}^r c_i w_i = 0.$$

Mas w_1, \dots, w_r são independentes e por isso $c_1 = \dots = c_r = 0$. Daqui resulta que os primeiros termos em (2.18) são zero, pelo que (2.18) se reduz a

$$\sum_{i=r+1}^{r+k} c_i e_i = 0.$$

Mas e_{r+1}, \dots, e_{r+k} são independentes visto formarem uma base para $N(T)$, e por isso $c_{r+1} = \dots = c_{r+k} = 0$. Porque todos os c_i em (2.18) são nulos, os elementos de (2.17) formam uma base para V , e o teorema está demonstrado.

EXEMPLO. Consideremos o Exemplo 2 da Secção 2.10, onde D é o operador derivação que aplica o espaço V dos polinómios de grau ≤ 3 no espaço W dos polinómios de grau ≤ 2 . Neste exemplo, o contradomínio $T(V) = W$, pelo que T tem ordem 3. Aplicando o método usado para demonstrar o teorema 2.14 definimos uma base qualquer para W , por exemplo a base $(1, x, x^2)$. Um conjunto de polinómios de V que se aplica nestes elementos é $(x, \frac{1}{2}x^2, \frac{1}{3}x^3)$. Amplia-se este conjunto para obtermos uma base para V juntando-lhe o polinómio constante 1, o qual é uma base para o espaço nulo de D . Deste modo, se utilizamos a base $(x, \frac{1}{2}x^2, \frac{1}{3}x^3, 1)$ para V e a base $(1, x, x^2)$ para W , a correspondente representação matricial para D tem a forma diagonal

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

2.12. Exercícios

Em todos os exercícios em que intervenha o espaço vectorial V_n , considera-se a base usual formada pelos vectores coordenados unitários, a menos que se diga expressamente o contrário. Nos exercícios relativos à matriz de uma transformação linear $T: V \rightarrow W$ com $V = W$, toma-se a mesma base quer em V quer em W , a menos que seja indicada outra escolha.

- Determinar a matriz de cada uma das seguintes transformações lineares de V_n em V_n :
 - a transformação identidade.
 - a transformação zero.
 - multiplicação por um escalar dado c .
- Determinar a matriz de cada uma das seguintes projecções.
 - $T: V_3 \rightarrow V_2$, onde $T(x_1, x_2, x_3) = (x_1, x_2)$.
 - $T: V_3 \rightarrow V_2$, onde $T(x_1, x_2, x_3) = (x_2, x_3)$.
 - $T: V_5 \rightarrow V_2$, onde $T(x_1, x_2, x_3, x_4, x_5) = (x_2, x_3, x_4)$.
- Uma transformação linear $T: V_2 \rightarrow V_2$ aplica os vectores da base i e j da maneira seguinte:

$$T(i) = i + j, \quad T(j) = 2i - j.$$

- Calcular $T(3i - 4j)$ e $T^2(3i - 4j)$ em função de i e j .
 - Determinar as matrizes de T e T^2 .
 - Resolver a alínea (b) se a base (i, j) é substituída por (e_1, e_2) , onde $e_1 = i - j$, $e_2 = 3i + j$.
- Uma transformação linear $T: V_2 \rightarrow V_2$ define-se do modo seguinte: Cada vector (x, y) transforma-se no seu simétrico relativamente ao eixo OY e depois duplica-se o seu comprimento para se obter $T(x, y)$. Determinar a matriz de T e de T^2 .
 - Seja $T: V_3 \rightarrow V_3$ uma transformação linear tal que

$$T(k) = 2i + 3j + 5k, \quad T(j + k) = i, \quad T(i + j + k) = j - k.$$

- (a) Calcular $T(i + 2j + 3k)$ e determinar a nulidade e ordem de T .
- (b) Determinar a matriz de T .
6. Para a transformação linear do Exercício 5, escolham-se ambas as bases definidas por (e_1, e_2, e_3) , onde $e_1 = (2, 3, 5)$, $e_2 = (1, 0, 0)$, $e_3 = (0, 1, -1)$; determinar a matriz de T relativa às novas bases.
7. Uma transformação linear $T: V_3 \rightarrow V_2$ aplica os vectores da base da maneira seguinte: $T(i) = (0, 0)$, $T(j) = (1, 1)$, $T(k) = (1, -1)$.
 - (a) Calcular $T(4i - j + k)$ e determinar a nulidade e a ordem de T .
 - (b) Determinar a matriz de T .
 - (c) Usar a base (i, j, k) de V_3 e a base (w_1, w_2) de V_2 com $w_1 = (1, 1)$, $w_2 = (1, 2)$. Determinar a matriz de T relativa a estas bases.
 - (d) Determinar bases (e_1, e_2, e_3) de V_3 e (w_1, w_2) de V_2 , relativamente às quais a matriz de T tenha a forma diagonal.
8. Uma transformação linear $T: V_2 \rightarrow V_3$ aplica os vectores da base do modo seguinte: $T(i) = (1, 0, 1)$, $T(j) = (-1, 0, 1)$.
 - (a) Calcular $T(2i - 3j)$ e determinar a nulidade (dimensão do núcleo) e a ordem T .
 - (b) Determinar a matriz de T .
 - (c) Achar bases (e_1, e_2) para V_2 e (w_1, w_2, w_3) para V_3 , para as quais a matriz de T tem a forma diagonal.
9. Resolver o Exercício 8 se $T(i) = (1, 0, 1)$ e $T(j) = (1, 1, 1)$.
10. Sejam V e W espaços lineares, cada um com dimensão 2 e ambos com a base (e_1, e_2) . Seja $T: V \rightarrow W$ uma transformação linear tal que

$$T(e_1 + e_2) = 3e_1 + 9e_2, \quad T(3e_1 + 2e_2) = 7e_1 + 23e_2.$$

- (a) Calcular $T(e_2 - e_1)$ e determinar a nulidade e a ordem de T .
- (b) Determinar a matriz de T relativa a uma base dada.
- (c) Utilizar a base (e_1, e_2) para V e determinar uma nova base da forma $(e_1 + ae_2, 2e_1 + be_2)$ para W , relativamente à qual a matriz de T terá a forma diagonal.

No espaço linear de todas as funções reais, cada um dos conjuntos seguintes é independente e gera um subespaço V de dimensão finita. Utilizar o conjunto dado como base para V e seja $D: V \rightarrow V$ o operador derivação. Em cada caso determinar a matriz de D e de D^2 , relativa à base que se indica.

- | | |
|---------------------------------|--|
| 11. $(\sin x, \cos x)$. | 15. $(-\cos x, \sin x)$. |
| 12. $(1, x, e^x)$. | 16. $(\sin x, \cos x, x \sin x, x \cos x)$. |
| 13. $(1, 1 + x, 1 + x + e^x)$. | 17. $(e^x \sin x, e^x \cos x)$. |
| 14. (e^x, xe^x) . | 18. $(e^{2x} \sin 3x, e^{2x} \cos 3x)$. |

19. Escolher a base $(1, x, x^2, x^3)$ no espaço linear V de todos os polinómios reais de grau ≤ 3 . Representando D o operador derivação, seja $T: V \rightarrow V$ a transformação linear que aplica $p(x)$ em $xp'(x)$. Relativamente à base dada, determinar a matriz de cada uma das seguintes transformações: (a) T ; (b) DT ; (c) TD ; (d) $TD - DT$; (e) T^2 ; (f) $T^2D^2 - D^2T^2$.
20. Considerar o Exercício 19. Seja W a imagem de V pela transformação TD . Determinar bases para V e para W relativamente às quais a matriz de TD tem a forma diagonal.

2.13. Espaços lineares de matrizes

Vimos como as matrizes se apresentam de uma maneira natural, como representações de transformações lineares. Mas as matrizes podem também considerar-se como existentes por direito próprio, sem necessariamente estarem ligados às transformações lineares. Consideradas como tal, formam outra classe de objectos matemáticos relativamente aos quais podem definir-se operações algébricas. A ligação com as transformações lineares serve como motivação para estas definições, mas esta ligação será ignorada por agora.

Sejam m e n dois inteiros positivos e $I_{m,n}$ o conjunto de todos os pares de inteiros (i, j) tal que $1 \leq i \leq m$, $1 \leq j \leq n$. Qualquer função A cujo domínio é $I_{m,n}$ chama-se uma *matriz* $m \times n$. O valor da função $A(i, j)$ chama-se o *elemento* ij da matriz e representa-se por a_{ij} . Habitualmente dispõem-se todos os valores da função num rectângulo por intermédio de m linhas e n colunas, como se indica

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

Os elementos a_{ij} podem ser objectos arbitrários de natureza qualquer. Usualmente serão números reais ou complexos, mas por vezes é conveniente considerar matrizes cujos elementos são de outra natureza, por exemplo funções. Também podemos representar as matrizes na notação mais compacta

$$A = (a_{ij})_{i,j=1}^{m,n} \quad \text{ou} \quad A = (a_{ij}).$$

Se $m = n$, a matriz diz-se *quadrada*. Uma matriz $1 \times n$ diz-se uma *matriz linha* e uma matriz $m \times 1$ diz-se uma *matriz coluna*.

Duas funções são iguais se e só se tiverem o mesmo domínio e tomarem os mesmos valores em cada elemento do domínio. Visto que as matrizes são funções, duas matrizes $A = (a_{ij})$ e $B = (b_{ij})$ são iguais se e só se tiverem o mesmo número de linhas e de colunas, e forem iguais os elementos $a_{ij} = b_{ij}$ para todo o par (i, j) .

Supondo agora que os elementos da matriz são números (reais ou complexos), vamos definir a adição de matrizes e a multiplicação por escalares pelo mesmo método usado para funções reais ou complexas quaisquer.

DEFINIÇÃO. Se $A = (a_{ij})$ e $B = (b_{ij})$ são duas matrizes $m \times n$ e se c é um escalar qualquer, definem-se as matrizes $A + B$ e cA do modo seguinte:

$$A + B = (a_{ij} + b_{ij}), \quad cA = (ca_{ij}).$$

A soma define-se unicamente quando A e B são do mesmo tipo $m \times n$ (mesmo número de linhas e mesmo número de colunas).

EXEMPLO. Se

$$A = \begin{bmatrix} 1 & 2 & -3 \\ -1 & 0 & 4 \end{bmatrix} \quad \text{e} \quad B = \begin{bmatrix} 5 & 0 & 1 \\ 1 & -2 & 3 \end{bmatrix},$$

temos pois

$$A + B = \begin{bmatrix} 6 & 2 & -2 \\ 0 & -2 & 7 \end{bmatrix}, \quad 2A = \begin{bmatrix} 2 & 4 & -6 \\ -2 & 0 & 8 \end{bmatrix}, \quad (-1)B = \begin{bmatrix} -5 & 0 & -1 \\ -1 & 2 & -3 \end{bmatrix}.$$

Definimos a *matriz nula* O , como sendo a matriz $m \times n$ na qual todos os elementos são 0. Com estas definições é um exercício simples verificar que o conjunto de todas as matrizes $m \times n$ define um espaço linear. Representamos este espaço linear por $M_{m,n}$. Se os elementos são números reais, o espaço $M_{m,n}$ é um espaço linear real. Se os elementos são complexos, $M_{m,n}$ é um espaço linear complexo. É igualmente fácil provar que este espaço tem dimensão mn . Com efeito, uma base para $M_{m,n}$ consiste de mn matrizes, tendo cada uma delas um elemento igual a 1 e todos os outros iguais a 0. Por exemplo, as seis matrizes

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

formam uma base para o conjunto de todas as matrizes 2×3 .

2.14. Isomorfismo entre transformações lineares e matrizes

Voltamos agora à relação entre matrizes e transformações lineares. Sejam V e W espaços lineares de dimensão finita, com $\dim V = n$ e $\dim W = m$. Escolhamos uma base (e_1, \dots, e_n) para V e uma base (w_1, \dots, w_m) para W . Nesta discussão estas bases consideram-se fixas. Seja $\mathcal{L}(V, W)$ o espaço linear de todas as transformações lineares de V em W . Se $T \in \mathcal{L}(V, W)$, seja $m(T)$ a matriz de T relativamente às bases dadas. Lembramos que $m(T)$ se define como segue:

A imagem de cada elemento base e_k exprime-se como uma combinação linear dos elementos da base de W :

$$(2.19) \quad T(e_k) = \sum_{i=1}^m t_{ik} w_i \quad \text{para} \quad k = 1, 2, \dots, n.$$

Os coeficientes escalares t_{ik} são os elementos ik de $m(T)$. Assim temos

$$(2.20) \quad m(T) = (t_{ik})_{i,k=1}^{m,n}.$$

A equação (2.20) define uma nova função m cujo domínio é $\mathcal{L}(V, W)$ e cujos valores são matrizes de $M_{m,n}$. Uma vez que cada matriz $m \times n$ é a matriz $m(T)$ para algum T em $\mathcal{L}(V, W)$, o contradomínio de m é $M_{m,n}$. O teorema que apresentamos a seguir mostra que a transformação $m: \mathcal{L}(V, W) \rightarrow M_{m,n}$ é linear e biunívoca em $\mathcal{L}(V, W)$.

TEOREMA 2.15. TEOREMA DO ISOMORFISMO. *Para todo S e todo T em $\mathcal{L}(V, W)$ e todo escalar c , tem-se*

$$m(S + T) = m(S) + m(T) \quad \text{e} \quad m(cT) = cm(T).$$

Além disso,

$$m(S) = m(T) \quad \text{implica} \quad S = T,$$

pelo que m é biunívoca em $\mathcal{L}(V, W)$.

Demonstração. A matriz $m(T)$ é formada pelos coeficientes t_{ik} de (2.19). Analogamente, a matriz $m(S)$ é formada pelos coeficientes s_{ik} nas equações

$$(2.21) \quad S(e_k) = \sum_{i=1}^m s_{ik} w_i \quad \text{para} \quad k = 1, 2, \dots, n.$$

Uma vez que se tem

$$(S + T)(e_k) = \sum_{i=1}^m (s_{ik} + t_{ik}) w_i \quad \text{e} \quad (cT)(e_k) = \sum_{i=1}^m (ct_{ik}) w_i,$$

obtemos $m(S + T) = (s_{ik} + t_{ik}) = m(S) + m(T)$ e $m(cT) = (ct_{ik}) = cm(T)$, o que provoca que m é linear.

Para provar que m é biunívoca, suponhamos que $m(S) = m(T)$, onde $S = (s_{ik})$ e $T = (t_{ik})$. As equações (2.19) e (2.21) mostram que $S(e_k) = T(e_k)$ para cada elemento da base e_k , pelo que $S(x) = T(x)$ para todo x de V , e por conseguinte $S = T$.

Nota: A função m chama-se um *isomorfismo*. Para uma dada escolha das bases, m estabelece uma correspondência biunívoca entre o conjunto de transformações lineares, $\mathcal{L}(V, W)$ e o conjunto $M_{m,n}$ de matrizes $m \times n$. As operações de adição e multiplicação por escalares conservam-se através desta correspondência. Os espaços lineares $\mathcal{L}(V, W)$ e $M_{m,n}$ dizem-se *isomorfos*. Incidentalmente, o teorema 2.11 mostra que o domínio de uma transformação linear biunívoca tem a mesma dimensão que o respectivo contradomínio. Portanto, $\dim \mathcal{L}(V, W) = \dim M_{m,n} = mn$.

Se $V = W$ e se escolhermos a mesma base em V e W , então a matriz $m(I)$ que corresponde à transformação identidade $I: V \rightarrow V$ é uma matriz diagonal $n \times n$, com cada elemento da diagonal igual à unidade e todos os restantes iguais a 0. Esta é a *matriz identidade* ou *matriz unidade* e representa-se por I ou por I_n .

2.15. Multiplicação de matrizes

Algumas transformações lineares podem multiplicar-se por meio da composição. Vamos passar a definir multiplicação de matrizes de tal maneira que o produto de duas matrizes corresponda à composição das transformações lineares que representam.

Recordamos que se $T: U \rightarrow V$ e $S: V \rightarrow W$ são transformações lineares, a sua composição $ST: U \rightarrow W$ é uma transformação linear dada por

$$ST(x) = S[T(x)] \text{ para todo } x \text{ de } U.$$

Suponhamos que U , V e W são de dimensão finita, por exemplo

$$\dim U = n, \quad \dim V = p, \quad \dim W = m.$$

Escolhamos bases para U , V e W . Relativamente a estas bases, a matriz $m(S)$ é uma matriz $m \times p$, a matriz T é uma matriz $p \times n$ e a matriz de ST é uma matriz $m \times n$. A definição que a seguir se apresenta de multiplicação de matrizes permite-nos deduzir a relação $m(ST) = m(S)m(T)$, o que estende aos produtos a propriedade de isomorfismo.

DEFINIÇÃO. *Sejam A uma matriz $m \times p$ qualquer, e B outra matriz $p \times n$ qualquer, por exemplo*

$$A = (a_{ij})_{i,j=1}^{m,p} \quad \text{e} \quad B = (b_{ij})_{i,j=1}^{p,n}.$$

O produto AB define-se como sendo a matriz $m \times n$, $C = (c_{ij})$, cujo elemento ij é dado por

$$(2.22) \quad c_{ij} = \sum_{k=1}^p a_{ik} b_{kj}.$$

Nota: O produto AB só se define quando o número de colunas de A é igual ao número de linhas de B .

Se escrevermos A_i para a linha de ordem i de A , e B^j para a coluna de ordem j de B e imaginamos estas como vectores p -dimensionais, então a soma (2.22) é simplesmente o produto escalar $A_i \cdot B^j$, isto é, o elemento ij de AB é o produto escalar da linha i de A com a coluna j de B :

$$AB = (A_i \cdot B^j)_{i,j=1}^{m,n}.$$

Assim a multiplicação de matrizes pode considerar-se como uma generalização de produto escalar.

EXEMPLO 1. Seja $A = \begin{bmatrix} 3 & 1 & 2 \\ -1 & 1 & 0 \end{bmatrix}$ e $B = \begin{bmatrix} 4 & 6 \\ 5 & -1 \\ 0 & 2 \end{bmatrix}$. Visto que A é 2×3 e B é 3×2 , o produto $A \cdot B$ é uma matriz 2×2 ,

$$AB = \begin{bmatrix} A_1 \cdot B^1 & A_1 \cdot B^2 \\ A_2 \cdot B^1 & A_2 \cdot B^2 \end{bmatrix} = \begin{bmatrix} 17 & 21 \\ 1 & -7 \end{bmatrix}.$$

Os elementos de A e B calculam-se do modo seguinte:

$$\begin{aligned} A_1 \cdot B^1 &= 3 \cdot 4 + 1 \cdot 5 + 2 \cdot 0 = 17, & A_1 \cdot B^2 &= 3 \cdot 6 + 1 \cdot (-1) + 2 \cdot 2 = 21, \\ A_2 \cdot B^1 &= (-1) \cdot 4 + 1 \cdot 5 + 0 \cdot 0 = 1, & A_2 \cdot B^2 &= (-1) \cdot 6 + 1 \cdot (-1) + 0 \cdot 2 = -7. \end{aligned}$$

EXEMPLO 2. Seja

$$A = \begin{bmatrix} 2 & 1 & -3 \\ 1 & 2 & 4 \end{bmatrix} \quad \text{e} \quad B = \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix}.$$

Aqui A é 2×3 e B é 3×1 , pelo que AB é a matriz 2×1 dada por

$$AB = \begin{bmatrix} A_1 \cdot B^1 \\ A_2 \cdot B^1 \end{bmatrix} = \begin{bmatrix} -9 \\ 8 \end{bmatrix},$$

visto que $A_1 \cdot B^1 = 2 \cdot (-2) + 1 \cdot 1 + (-3) \cdot 2 = -9$ e $A_2 \cdot B^1 = 1 \cdot (-2) + 2 \cdot 1 + 4 \cdot 2 = 8$.

EXEMPLO 3. Se A e B são ambas matrizes quadradas da mesma ordem, então define-se AB e BA . Por exemplo, se

encontramos que

$$A = \begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix} \quad \text{e} \quad B = \begin{bmatrix} 3 & 4 \\ 5 & 2 \end{bmatrix},$$

$$AB = \begin{bmatrix} 13 & 8 \\ 2 & -2 \end{bmatrix}, \quad BA = \begin{bmatrix} -1 & 10 \\ 3 & 12 \end{bmatrix}.$$

Este exemplo prova que $AB \neq BA$, o que acontece em geral. Se $AB = BA$, dizemos que A e B *comutam* (ou que são *permutáveis*).

EXEMPLO 4. Se I_p é a matriz identidade $p \times p$, então $I_p A = A$ para qualquer matriz A , $p \times n$, e $B I_p = B$ para qualquer B , $m \times p$. Por exemplo,

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}.$$

Provamos agora que a matriz da composição ST é o produto de matrizes $m(S)$ e $m(T)$.

TEOREMA 2.16. *Se $T: U \rightarrow V$ e $S: V \rightarrow W$ são transformações lineares com U, V e W espaços lineares de dimensão finita, então, para determinada escolha das bases, as matrizes de S, T e ST estão relacionadas por*

$$m(ST) = m(S)m(T).$$

Demonstração. Suponhamos que $\dim U = n$, $\dim V = p$, $\dim W = m$. Sejam (u_1, \dots, u_n) uma base para U , (v_1, \dots, v_p) uma base para V , e (w_1, \dots, w_m) uma base para W . Relativamente a estas bases, temos

$$m(S) = (s_{ij})_{i,j=1}^{m,p}, \quad \text{onde} \quad S(v_k) = \sum_{i=1}^m s_{ik} w_i \quad \text{para} \quad k = 1, 2, \dots, p,$$

e

$$m(T) = (t_{ij})_{i,j=1}^{p,n}, \quad \text{onde} \quad T(u_j) = \sum_{k=1}^p t_{kj} v_k \quad \text{para} \quad j = 1, 2, \dots, n.$$

Desta maneira, temos

$$ST(u_j) = S[T(u_j)] = \sum_{k=1}^p t_{kj} S(v_k) = \sum_{k=1}^p t_{kj} \sum_{i=1}^m s_{ik} w_i = \sum_{i=1}^m \left(\sum_{k=1}^p s_{ik} t_{kj} \right) w_i,$$

pelo que se encontra

$$m(ST) = \left(\sum_{k=1}^p s_{ik} t_{kj} \right)_{i,j=1}^{m,n} = m(S)m(T).$$

Já se observou que a multiplicação de matrizes nem sempre satisfaz à propriedade comutativa. O teorema seguinte mostra, porém, que a operação goza de propriedade associativa e distributiva.

TEOREMA 2.17. PROPRIEDADES ASSOCIATIVA E DISTRIBUTIVA PARA A MULTIPLICAÇÃO DE MATRIZES. *Dadas as matrizes A, B, C .*

(a) *Se os produtos $A(BC)$ e $(AB)C$ têm sentido, tem-se*

$$A(BC) = (AB)C \quad (\text{propriedade associativa}).$$

(b) *Suponha-se que A e B são do mesmo tipo. Se AC e BC têm significado, tem-se*

$$(A + B)C = AC + BC \quad (\text{propriedade distributiva, à direita}),$$

enquanto que se CA e CB têm significado, se tem

$$C(A + B) = CA + CB \quad (\text{propriedade distributiva, à esquerda}).$$

Demonstração. Estas propriedades podem deduzir-se directamente da definição do produto de matrizes, mas preferimos o seguinte tipo de raciocínio. Introduzimos espaços lineares de dimensão finita U, V, W, X e transformações lineares $T: U \rightarrow V, S: V \rightarrow W, R: W \rightarrow X$ tais que, para uma escolha fixa das bases, se tem

$$A = m(R), \quad B = m(S), \quad C = m(T).$$

Pelo teorema 2.16, temos $m(RS) = AB$ e $m(ST) = BC$. Da propriedade associativa para a composição, encontramos que $R(ST) = (RS)T$. Aplicando o teorema 2.16 uma vez mais a esta igualdade, obtemos $m(R)m(ST) = m(RS)m(T)$, ou $A(BC) = (AB)C$, o que demonstra (a). A demonstração de (b) pode fazer-se por um tipo de raciocínio análogo.

DEFINIÇÃO. Se A é uma matriz quadrada, definem-se potências inteiras de A , por indução, do modo seguinte:

$$A^0 = I, \quad A^n = AA^{n-1} \quad \text{for } n \geq 1.$$

2.16. Exercícios

$$1. \text{ Se } A = \begin{bmatrix} 1 & -4 & 2 \\ -1 & 4 & -2 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 \\ -1 & 3 \\ 5 & -2 \end{bmatrix}, \quad C = \begin{bmatrix} 2 & 2 \\ 1 & -1 \\ 1 & -3 \end{bmatrix}, \text{ calcular } B + C, AB,$$

$$BA, AC, CA, A(2B - 3C).$$

$$2. \text{ Seja } A = \begin{bmatrix} 0 & 1 \\ 0 & 2 \end{bmatrix}. \text{ Determinar todas as matrizes } B, 2 \times 2, \text{ tais que (a) } AB = O; \text{ (b) } BA = O.$$

3. Para cada alínea determina a, b, c, d que satisfaçam à equação dada

$$(a) \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} 1 \\ 9 \\ 6 \\ 5 \end{bmatrix}; \quad (b) \begin{bmatrix} a & b & c & d \\ 1 & 4 & 9 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 6 & 6 \\ 1 & 9 & 8 & 4 \end{bmatrix}.$$

4. Calcular, para cada alínea, $AB - BA$.

$$(a) A = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 1 & 2 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 4 & 1 & 1 \\ -4 & 2 & 0 \\ 1 & 2 & 1 \end{bmatrix};$$

$$(b) A = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 1 & 2 \\ -1 & 2 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & 1 & -2 \\ 3 & -2 & 4 \\ -3 & 5 & 11 \end{bmatrix}.$$

5. Se A é uma matriz quadrada, provar que $A^n A^m = A^{m+n}$, quaisquer que sejam os inteiros $m \geq 0$ e $n \geq 0$.

6. Seja $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Verificar que $A^2 = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$ e calcular A^n .

7. Seja $A = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$. Verificar que $A^2 = \begin{bmatrix} \cos 2\theta & -\sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{bmatrix}$ e calcular A^n .

8. Seja $A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$. Verificar que $A^2 = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$. Calcular A^3 e A^4 . Tentar escrever

uma fórmula geral para A^n e demonstrá-la por indução.

9. Seja $A = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$. Provar que $A^2 = 2A - I$ e calcular A^{100} .

10. Determinar todas as matrizes A , 2×2 , tais que $A^2 = O$.

11. (a) Provar que uma matriz A , 2×2 , comuta com toda a matriz 2×2 se e só se A comuta com cada uma das quatro matrizes

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

(b) Achar todas as matrizes A .

12. A equação $A^2 = I$ é satisfeita por cada uma das matrizes 2×2

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ c & -1 \end{bmatrix}, \quad \begin{bmatrix} 1 & b \\ 0 & -1 \end{bmatrix},$$

onde b e c são números reais arbitrários. Determinar todas as matrizes A , 2×2 , tais que $A^2 = I$.

13. Se $A = \begin{bmatrix} 2 & -1 \\ -2 & 3 \end{bmatrix}$ e $B = \begin{bmatrix} 7 & 6 \\ 9 & 8 \end{bmatrix}$, achar as matrizes C e D , 2×2 , tais que $AC = B$ e $DA = B$.

14. (a) Verificar que as identidades algébricas

$$(A + B)^2 = A^2 + 2AB + B^2 \quad \text{e} \quad (A + B)(A - B) = A^2 - B^2$$

não são válidas para as matrizes 2×2 , $A = \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix}$ e $B = \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}$.

(b) Modificar os segundos membros daquelas identidades para obter fórmulas válidas para todas as matrizes quadradas A e B .

(c) Para que matrizes A e B são válidas as identidades dadas em (a)?

2.17. Sistemas de equações lineares

Seja $A = (a_{ij})$ uma dada matriz $m \times n$ de números, e sejam c_1, c_2, \dots, c_m outros m números dados. Um conjunto de m equações da forma

$$(2.23) \quad \sum_{k=1}^n a_{ik}x_k = c_i \quad \text{para } i = 1, 2, \dots, m,$$

chama-se um sistema de m equações lineares com n incógnitas. Aqui x_1, x_2, \dots, x_n representam as incógnitas. Uma *solução* do sistema é qualquer sistema de n números (x_1, \dots, x_n) para os quais todas as equações sejam satisfeitas. A matriz A chama-se a *matriz dos coeficientes das incógnitas do sistema*.

Sistemas lineares podem estudar-se com auxílio das transformações lineares. Escolhamos as bases usuais de vectores coordenados unitários em V_n e V_m . A matriz dos coeficientes A determina uma transformação linear, $T: V_n \rightarrow V_m$, a qual aplica um vector arbitrário $x = (x_1, \dots, x_n)$ de V_n num vector $y = (y_1, \dots, y_m)$ de V_m , definido pelas m equações lineares

$$y_i = \sum_{k=1}^n a_{ik}x_k \quad \text{para } i = 1, 2, \dots, m.$$

Seja $c = (c_1, \dots, c_m)$ o vector de V_m cujas componentes são os números aparecendo no sistema (2.23). Este sistema pode escrever-se mais simplesmente

$$T(x) = c.$$

O sistema admite uma solução se e só se c pertence ao contradomínio de T . Se um só x de V_n se aplica em c , o sistema tem uma solução única. Se mais do que um x se aplica em c , o sistema tem mais do que uma solução.

EXEMPLO 1. *Um sistema sem solução.* O sistema $x + y = 1$, $x + y = 2$ não admite solução. A soma de dois números não pode ser simultaneamente 1 e 2.

EXEMPLO 2. *Um sistema com solução única.* O sistema $x + y = 1$, $x - y = 0$ tem precisamente uma solução: $(x, y) = (\frac{1}{2}, \frac{1}{2})$.

EXEMPLO 3. *Um sistema com mais de uma solução.* O sistema $x + y = 1$, formado por uma única equação com duas incógnitas, tem mais do que uma solução. Quaisquer dois números cuja soma seja 1 definem uma solução do sistema.

A cada sistema linear (2.23), podemos associar outro sistema

$$\sum_{k=1}^n a_{ik}x_k = 0 \quad \text{para } i = 1, 2, \dots, m,$$

obtido pela substituição de cada c_i , em (2.23), por 0. Este diz-se o *sistema homogêneo* correspondente a (2.23). Se $c \neq 0$, o sistema (2.23) diz-se um sistema *não homogêneo*. Um vector x de V_n verificará o sistema homogêneo se e só se

$$T(x) = 0,$$

onde T é uma transformação linear definida pela matriz dos coeficientes. O sistema homogêneo tem sempre a solução, $x = 0$, mas pode ter outras. O conjunto de soluções do sistema homogêneo é o espaço nulo de T . O teorema que se apresenta a seguir descreve a relação entre as soluções de um sistema não homogêneo e as do sistema homogêneo correspondente.

TEOREMA 2.18. *Admite-se que o sistema não homogêneo (2.23) tem uma solução, por exemplo b .*

- (a) *Se um vector x é uma solução do sistema não homogêneo, então o vector $v = x - b$ é uma solução do correspondente sistema homogêneo.*
- (b) *Se um vector v é uma solução do sistema homogêneo, então o vector $x = v + b$ é uma solução do sistema não homogêneo.*

Demonstração. Seja $T: V_n \rightarrow V_m$ uma transformação linear definida pela matriz dos coeficientes, como se referiu atrás. Uma vez que b é uma solução do sistema não homogêneo, temos $T(b) = c$. Sejam x e v dois vectores de V_n tais que $v = x - b$. Então tem-se

$$T(v) = T(x - b) = T(x) - T(b) = T(x) - c.$$

Portanto $T(x) = c$ se e só se $T(v) = 0$, o que demonstra simultaneamente (a) e (b).

Este teorema mostra que o problema da determinação de todas as soluções de um sistema não homogêneo se divide naturalmente em duas partes: (1) Determinação de todas as soluções v do sistema homogêneo, isto é, determinação do espaço nulo de T , e (2) determinação de uma solução particular b do sistema não homogêneo. Adicionando b a cada vector v do espaço nulo de T , obtemos todas as soluções $x = v + b$ do sistema não homogêneo.

Seja k a dimensão de $N(T)$ (a nulidade de T). Se pudermos encontrar k soluções independentes v_1, \dots, v_k do sistema homogêneo, elas formarão uma base para $N(T)$, e podemos obter cada v em $N(T)$ formando todas as combinações lineares possíveis.

$$v = t_1 v_1 + \cdots + t_k v_k,$$

onde t_1, \dots, t_k são escalares arbitrários. Esta combinação linear chama-se a *solução do sistema homogéneo*. Se b é uma solução particular do sistema não homogéneo, então todas as soluções x são dadas por

$$x = b + t_1 v_1 + \cdots + t_k v_k.$$

Esta combinação linear diz-se a *solução geral do sistema não homogéneo*. Ao teorema 2.18 pode então dar-se a nova forma.

TEOREMA 2.19. *Seja $T: V_n \rightarrow V_m$ uma transformação tal que $T(x) = y$, onde $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_m)$ e*

$$y_i = \sum_{k=1}^n a_{ik} x_k \quad \text{para } i = 1, 2, \dots, m.$$

Seja k a nulidade de T . Se v_1, \dots, v_k são k soluções independentes do sistema homogéneo $T(x) = 0$, e se b é uma solução particular do sistema não homogéneo $T(x) = c$, então a solução geral do sistema não homogéneo é

$$x = b + t_1 v_1 + \cdots + t_k v_k,$$

onde t_1, \dots, t_k são escalares arbitrários.

Este teorema nada nos diz como decidir se um sistema não homogéneo tem uma solução particular b , nem nos diz como determinar as soluções v_1, \dots, v_k do sistema homogéneo. Diz-nos apenas o que pode obter-se quando o sistema não homogéneo tem uma solução. O exemplo seguinte, embora muito simples, ilustra o teorema.

EXEMPLO. O sistema $x + y = 2$ admite como sistema homogéneo associado, $x + y = 0$. Por tal facto, o espaço nulo é formado por todos os vectores de V_2 da forma $(t, -t)$ com t arbitrário. Porque $(t, -t) = t(1, -1)$, este é um subespaço unidimensional de V_2 com base $(1, -1)$. Uma solução particular do sistema não homogéneo é $(0, 2)$. Por conseguinte, a solução geral do sistema não homogéneo é dada por

$$(x, y) = (0, 2) + t(1, -1) \quad \text{ou} \quad x = t, \quad y = 2 - t,$$

com t arbitrário.

2.18. Técnicas de cálculo

Voltamos ao problema do cálculo de soluções de um sistema linear não homogéneo. Embora tenham sido desenvolvidos muitos métodos para resolver este problema,

todos eles requerem cálculos consideráveis se o sistema tem muitas equações. Por exemplo, para resolvermos um sistema de dez equações com o mesmo número de incógnitas podemos levar várias horas de cálculos, ainda que disponhamos de um calculador manual.

Vamos analisar um método largamente utilizado, conhecido por *método de eliminação de Gauss-Jordan*, o qual é relativamente simples e pode ser facilmente programado por calculadores electrónicos de alta velocidade. O método consiste na aplicação de três tipos de operações às equações do sistema linear.

- (1) *Troca, entre si, de duas equações;*
- (2) *Multiplicação de todos os termos de uma equação por um escalar não nulo;*
- (3) *Adição a uma equação, membro a membro, de outra multiplicada por um escalar.*

Cada vez que efectuamos uma destas operações no sistema, obtemos um novo sistema tendo precisamente as mesmas soluções que o anterior. Dois tais sistemas dizem-se *equivalentes*. Efectuando estas operações uma a seguir à outra, de uma maneira sistemática e conveniente, chegamos finalmente a um sistema equivalente que pode resolver-se por uma análise simples.

Ilustraremos o método com alguns exemplos concretos. Resultará então claro como deve o método ser aplicado em geral.

EXEMPLO 1. *Um sistema com uma única solução.* Consideremos o sistema

$$\begin{aligned} 2x - 5y + 4z &= -3 \\ x - 2y + z &= 5 \\ x - 4y + 6z &= 10. \end{aligned}$$

Este sistema tem a solução única, $x = 124$, $y = 75$, $z = 31$, que vamos obter pelo método de eliminação de Gauss-Jordan. Para evitar trabalho não copiamos as letras x , y , z nem os sinais de igualdade, mas pelo contrário trabalharemos com a *matriz ampliada*

$$(2.24) \quad \left[\begin{array}{ccc|c} 2 & -5 & 4 & -3 \\ 1 & -2 & 1 & 5 \\ 1 & -4 & 6 & 10 \end{array} \right]$$

obtida pela junção dos segundos membros das equações do sistema à matriz dos coeficientes. Os três tipos de operações fundamentais mencionados atrás efectuam-se com as linhas da matriz ampliada e chamam-se por esse facto, *operações linha*. Em qualquer fase do processo, para obtermos as equações, colocamos as letras x , y e z e intercalamos os sinais de igualdade ao longo da linha divisória da matriz. O nosso objectivo consiste em chegarmos a

$$(2.25) \quad \left[\begin{array}{ccc|c} 1 & 0 & 0 & 124 \\ 0 & 1 & 0 & 75 \\ 0 & 0 & 1 & 31 \end{array} \right]$$

a matriz ampliada depois de uma sucessão de operações linha. O correspondente sistema de equações é $x = 124$, $y = 75$, $z = 31$, que é a solução pretendida.

A primeira fase consiste em obter o elemento 1 no vértice superior esquerdo da matriz. Podemos consegui-lo trocando entre si as duas primeiras linhas da matriz (2.24), ou então multiplicando a primeira linha por $\frac{1}{2}$. Permutando as duas primeiras linhas obtemos

$$\left[\begin{array}{ccc|c} 1 & -2 & 1 & 5 \\ 2 & -5 & 4 & -3 \\ 1 & -4 & 6 & 10 \end{array} \right]$$

A fase seguinte consiste em tornar nulos todos os elementos da primeira coluna, deixando a primeira linha como está. Para se conseguir tal, multiplicamos a primeira linha por -2 e adicionamos o resultado à segunda linha; em seguida multiplicamos a primeira linha por -1 e adicionamos o resultado à terceira linha. Depois destas duas operações, obtemos

$$(2.26) \quad \left[\begin{array}{ccc|c} 1 & -2 & 1 & 5 \\ 0 & -1 & 2 & -13 \\ 0 & -2 & 5 & 5 \end{array} \right]$$

Repetimos agora o processo na matriz $\left[\begin{array}{cc|c} -1 & 2 & -13 \\ -2 & 5 & 5 \end{array} \right]$, que aparece adjacente a dois zeros. Podemos obter 1 no *seu* vértice superior esquerdo multiplicando a segunda linha de (2.26) por -1 . Obtemos então a matriz

$$\left[\begin{array}{ccc|c} 1 & -2 & 1 & 5 \\ 0 & 1 & -2 & 13 \\ 0 & -2 & 5 & 5 \end{array} \right]$$

Multiplicando a segunda linha por 2 e adicionando o resultado à terceira, obtemos.

$$(2.27) \quad \left[\begin{array}{ccc|c} 1 & -2 & 1 & 5 \\ 0 & 1 & -2 & 13 \\ 0 & 0 & 1 & 31 \end{array} \right]$$

Nesta fase o sistema de equações inicial pode escrever-se

$$x - 2y + z = 5$$

$$y - 2z = 13$$

$$z = 31.$$

Estas equações podem resolver-se sucessivamente, partindo da terceira e caminhando em sentido inverso, para obtermos

$$z = 31, \quad y = 13 + 2z = 13 + 62 = 75, \quad x = 5 + 2y - z = 5 + 150 - 31 = 124.$$

Ou então, podemos continuar o processo de Gauss-Jordan convertendo em zeros todos os elementos situados acima da primeira diagonal na segunda e terceira colunas. Multiplicando a segunda linha de (2.27) por 2 e adicionando o resultado à primeira linha, obtemos

$$\left[\begin{array}{ccc|c} 1 & 0 & -3 & 31 \\ 0 & 1 & -2 & 13 \\ 0 & 0 & 1 & 31 \end{array} \right].$$

Finalmente multiplicamos a terceira linha por 3 e adicionamos o resultado à primeira e em seguida multiplicamos a terceira linha por 2 e adicionamos o resultado à segunda para obtermos a matriz (2.25).

EXEMPLO 2. *Um sistema com mais do que uma solução.* Consideremos o seguinte sistema de 3 equações com 5 incógnitas:

$$(2.28) \quad \begin{aligned} 2x - 5y + 4z + u - v &= -3 \\ x - 2y + z - u + v &= 5 \\ x - 4y + 6z + 2u - v &= 10. \end{aligned}$$

A correspondente matriz ampliada é

$$\left[\begin{array}{ccccc|c} 2 & -5 & 4 & 1 & -1 & -3 \\ 1 & -2 & 1 & -1 & 1 & 5 \\ 1 & -4 & 6 & 2 & -1 & 10 \end{array} \right].$$

Os coeficientes de x , y , z e os segundos membros são os mesmos do Exemplo 1. Se efectuarmos a mesmas operações linha referidas no Exemplo 1, obtemos a matriz ampliada

$$\left[\begin{array}{ccccc|c} 1 & 0 & 0 & -16 & 19 & 124 \\ 0 & 1 & 0 & -9 & 11 & 75 \\ 0 & 0 & 1 & -3 & 4 & 31 \end{array} \right].$$

O correspondente sistema de equações pode resolver-se relativamente a x, y, z em função de u e v dando-nos

$$x = 124 + 16u - 19v$$

$$y = 75 + 9u - 11v$$

$$z = 31 + 3u - 4v.$$

Se fizemos $u = t_1$ e $v = t_2$, com t_1 e t_2 números reais arbitrários, e determinarmos x, y, z por estas equações, o vector (x, y, z, u, v) , em V_5 , dado por

$$(x, y, z, u, v) = (124 + 16t_1 - 19t_2, 75 + 9t_1 - 11t_2, 31 + 3t_1 - 4t_2, t_1, t_2)$$

é uma solução. Separando as partes contendo t_1 e t_2 , podemos escrever a igualdade anterior na forma

$$(x, y, z, u, v) = (124, 75, 31, 0, 0) + t_1(16, 9, 3, 1, 0) + t_2(-19, -11, -4, 0, 1).$$

Esta equação dá a solução geral do sistema. O vector $(124, 75, 31, 0, 0)$ é uma solução particular do sistema não homogéneo (2.28). Os dois vectores $(16, 9, 3, 1, 0)$ e $(-19, -11, -4, 0, 1)$ são soluções do correspondente sistema homogéneo. Uma vez que são independentes, constituem uma base para o espaço de todas as soluções do sistema homogéneo.

EXEMPLO 3. *Um sistema sem solução.* Consideremos o sistema

$$2x - 5y + 4z = -3$$

$$(2.29) \quad x - 2y + z = 5$$

$$x - 4y + 5z = 10.$$

Este sistema é quase idêntico ao do Exemplo 1, excepto que o coeficiente de z na terceira equação se mudou de 6 para 5. A correspondente matriz ampliada é

$$\left[\begin{array}{ccc|c} 2 & -5 & 4 & -3 \\ 1 & -2 & 1 & 5 \\ 1 & -4 & 5 & 10 \end{array} \right].$$

Efectuando as mesmas operações do Exemplo 1 sobre as linhas de matriz anterior que permitiram a passagem de (2.24) a (2.27), obtemos a matriz ampliada

$$(2.30) \quad \left[\begin{array}{ccc|c} 1 & -2 & 1 & 5 \\ 0 & 1 & -2 & 13 \\ 0 & 0 & 0 & 31 \end{array} \right].$$

Quando a última linha se escreve na forma de equação, exprime que $0 = 31$. Por conseguinte o sistema inicial não admite solução uma vez que os dois sistemas (2.29) e (2.30) são equivalentes.

Em cada um dos exemplos precedentes, o número de equações não excedia o número de incógnitas. Se o número de equações for superior ao das incógnitas, o método de Gauss-Jordan é ainda aplicável. Por exemplo, suponhamos o sistema do Exemplo 1, o qual têm a solução $x = 124$, $y = 75$, $z = 31$. Se juntamos uma nova equação a este sistema que seja satisfeita pelo mesmo termo de números, por exemplo a equação $2x - 3y + z = 54$, então o método de eliminação conduz à matriz ampliada

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 124 \\ 0 & 1 & 0 & 75 \\ 0 & 0 & 1 & 31 \\ 0 & 0 & 0 & 0 \end{array} \right]$$

com uma linha de zeros na parte inferior. Contudo se juntarmos ao sistema inicial uma nova equação que não seja satisfeita pelos números $(124, 75, 31)$, por exemplo a equação $x + y + z = 1$, então o método de eliminação conduz a uma matriz ampliada da forma

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 124 \\ 0 & 1 & 0 & 75 \\ 0 & 0 & 1 & 31 \\ 0 & 0 & 0 & a \end{array} \right],$$

com $a \neq 0$. A última linha dá-nos uma equação $0 = a$, a qual nos leva a concluir que o sistema não admite solução.

2.19. Inversas de matrizes quadradas

Seja $A = (a_{ij})$ uma matriz quadrada $n \times n$. Se existir outra matriz quadrada $n \times n$, B , tal que $BA = I$, com I a matriz identidade $n \times n$, então A diz-se *não singular* e B chama-se a *inversa esquerda* de A .

Escolhamos a base usual do sistema de vectores unitários coordenados de V_n e seja $T: V_n \rightarrow V_n$ uma transformação linear com matriz $m(T) = A$. Então temos o seguinte.

TEOREMA 2.20. *A matriz A é não singular (regular) se e só se T é invertível. Se $BA = I$, então $B = m(T^{-1})$.*

Demonstração. Admitamos que A é não singular e que $BA = I$. Vamos provar que $T(x) = O$ implica $x = O$. Dado x , tal que $T(x) = O$, seja X a matriz coluna $n \times 1$ formada pelas componentes de x . Uma vez que $T(x) = O$, a matriz produto AX é uma matriz coluna $n \times 1$ formada de zeros, pelo que $B(AX)$ é também uma matriz coluna de zeros. Mas $B(AX) = (BA)X = IX = X$, pelo que toda a componente de x é 0. Deste modo, T é invertível, e a equação $TT^{-1} = I$ implica que $m(T)m(T^{-1}) = I$ ou $Am(T^{-1}) = I$. Multiplicando à esquerda por B , encontramos $m(T^{-1}) = B$. Inversamente, se T é invertível, então $T^{-1}T$ é a transformação identidade pelo que $m(T^{-1})m(T)$ é a matriz identidade. Deste modo A é não singular e $m(T^{-1})A = I$.

Todas as propriedades das transformações lineares invertíveis tem equivalente nas matrizes não singulares. Em particular, matrizes inversas esquerdas (se existirem) são únicas e toda a inversa esquerda é também inversa direita. Por outras palavras, se A é não singular e $BA = I$, então $AB = I$. Chama-se B a *inversa* de A e representa-se por A^{-1} . A inversa A^{-1} é também não singular e a sua inversa é A .

Vamos agora demonstrar que o problema da determinação dos elementos da inversa de uma matriz não singular é equivalente à resolução de n sistemas diferentes de equações lineares não homogêneas.

Seja $A = (a_{ij})$ não singular e $A^{-1} = (b_{ij})$ a sua inversa. Os elementos de A e A^{-1} estão relacionados pelas n^2 equações

$$(2.31) \quad \sum_{k=1}^n a_{ik}b_{kj} = \delta_{ij},$$

onde $\delta_{ij} = 1$ se $i = j$, $\delta_{ij} = 0$ se $i \neq j$. Para cada valor fixo de j , pode considerar-se (2.31) um sistema não homogêneo de n equações lineares com n incógnitas $b_{1j}, b_{2j}, \dots, b_{nj}$. Uma vez que A é não singular, cada um destes sistemas tem uma solução única, a coluna j de B . Todos estes sistemas têm a mesma matriz de coeficientes A e diferem unicamente nos seus segundos membros. Por exemplo, se A é uma matriz 3×3 , existem 9 equações em (2.31) que podem exprimir-se como 3 sistemas distintos de equações lineares com as seguintes matrizes ampliadas.

$$\left[\begin{array}{ccc|c} a_{11} & a_{12} & a_{13} & 1 \\ a_{21} & a_{22} & a_{23} & 0 \\ a_{31} & a_{32} & a_{33} & 0 \end{array} \right], \quad \left[\begin{array}{ccc|c} a_{11} & a_{12} & a_{13} & 0 \\ a_{21} & a_{22} & a_{23} & 1 \\ a_{31} & a_{32} & a_{33} & 0 \end{array} \right], \quad \left[\begin{array}{ccc|c} a_{11} & a_{12} & a_{13} & 0 \\ a_{21} & a_{22} & a_{23} & 0 \\ a_{31} & a_{32} & a_{33} & 1 \end{array} \right].$$

Se aplicamos o método de Gauss-Jordan, chegamos às matrizes ampliadas

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & b_{11} \\ 0 & 1 & 0 & b_{21} \\ 0 & 0 & 1 & b_{31} \end{array} \right], \quad \left[\begin{array}{ccc|c} 1 & 0 & 0 & b_{12} \\ 0 & 1 & 0 & b_{22} \\ 0 & 0 & 1 & b_{32} \end{array} \right], \quad \left[\begin{array}{ccc|c} 1 & 0 & 0 & b_{13} \\ 0 & 1 & 0 & b_{23} \\ 0 & 0 & 1 & b_{33} \end{array} \right].$$

Na prática aproveita-se o facto de que os três sistemas têm a mesma matriz de coeficientes e resolvem-se os três sistemas de uma vez trabalhando com a matriz ampliada

$$\left[\begin{array}{ccc|ccc} a_{11} & a_{12} & a_{13} & 1 & 0 & 0 \\ a_{21} & a_{22} & a_{23} & 0 & 1 & 0 \\ a_{31} & a_{32} & a_{33} & 0 & 0 & 1 \end{array} \right].$$

O método de eliminação conduz-nos a

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & b_{11} & b_{12} & b_{13} \\ 0 & 1 & 0 & b_{21} & b_{22} & b_{23} \\ 0 & 0 & 1 & b_{31} & b_{32} & b_{33} \end{array} \right].$$

A matriz à direita da linha divisória é a matriz inversa que se pretendia obter, sendo a matriz à esquerda dessa linha a matriz identidade.

Não é necessário saber à priori se A é não singular. Se A é *singular* (não regular), podemos ainda aplicar o método de Gauss-Jordan, mas acontecerá que no processo de diagonalização alguns dos elementos da diagonal virão nulos, e não será possível transformar A na matriz identidade.

Um sistema de n equações lineares com n incógnitas está representado por

$$\sum_{k=1}^n a_{ik}x_k = c_i, \quad i = 1, 2, \dots, n,$$

pode-se escrever duma maneira mais simples como uma equação matricial.

$$AX = C,$$

onde $A = (a_{ij})$ é a matriz dos coeficientes, e X e C matrizes coluna,

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad C = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}.$$

Se A é não singular, a única solução do sistema vem dada por $X = A^{-1}C$

2.20. Exercícios.

Aplicando o método de Gauss-Jordan a cada um dos sistemas seguintes, determinar a solução geral caso exista

$$\begin{aligned} 1. \quad & x + y + 3z = 5 \\ & 2x - y + 4z = 11 \\ & -y + z = 3. \end{aligned}$$

$$\begin{aligned} 2. \quad & 3x + 2y + z = 1 \\ & 5x + 3y + 3z = 2 \\ & x + y - z = 1. \end{aligned}$$

$$\begin{aligned} 3. \quad & 3x + 2y + z = 1 \\ & 5x + 3y + 3z = 2 \\ & 7x + 4y + 5z = 3. \end{aligned}$$

$$\begin{aligned} 4. \quad & 3x + 2y + z = 1 \\ & 5x + 3y + 3z = 2 \\ & 7x + 4y + 5z = 3 \\ & x + y - z = 0. \end{aligned}$$

$$\begin{aligned} 5. \quad & 3x - 2y + 5z + u = 1 \\ & x + y - 3z + 2u = 2 \\ & 6x + y - 4z + 3u = 7. \end{aligned}$$

$$\begin{aligned} 6. \quad & x + y - 3z + u = 5 \\ & 2x - y + z - 2u = 2 \\ & 7x + y - 7z + 3u = 3. \end{aligned}$$

$$\begin{aligned} 7. \quad & x + y + 2z + 3u + 4v = 0 \\ & 2x + 2y + 7z + 11u + 14v = 0 \\ & 3x + 3y + 6z + 10u + 15v = 0. \end{aligned}$$

$$\begin{aligned} 8. \quad & x - 2y + z + 2u = -2 \\ & 2x + 3y - z - 5u = 9 \\ & 4x - y + z - u = 5 \\ & 5x - 3y + 2z + u = 3. \end{aligned}$$

9. Demonstrar que o sistema $x + y + 2z = 2$, $2x - y + 3z = 2$, $5x - y + az = 6$, tem uma solução única se $a \neq 8$. Determinar todas as soluções se $a = 8$.

10. (a) Determinar todas as soluções do sistema

$$\begin{aligned} 5x + 2y - 6z + 2u &= -1 \\ x - y + z - u &= -2. \end{aligned}$$

(b) Determinar todas as soluções do sistema

$$\begin{aligned} 5x + 2y - 6z + 2u &= -1 \\ x - y + z - u &= -2 \\ x + y + z &= 6. \end{aligned}$$

11. Este exercício mostra-nos como determinar todas as matrizes não singulares 2×2 . Provar que

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = (ad - bc)I.$$

Deduzir que $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ é não singular se e só se $ad - bc \neq 0$, caso em que a sua inversa é

$$\frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Determinar a inversa de cada uma das matrizes dos Exercícios 12 a 16.

$$12. \begin{bmatrix} -2 & 3 & 4 \\ 2 & 1 & 1 \\ -1 & 1 & 2 \end{bmatrix}.$$

$$15. \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

$$13. \begin{bmatrix} 1 & 2 & 2 \\ 2 & -1 & 1 \\ 1 & 3 & 2 \end{bmatrix}.$$

$$16. \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 2 & 0 & 2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 & 0 & 1 \\ 0 & 0 & 0 & 0 & 2 & 0 \end{bmatrix}.$$

$$14. \begin{bmatrix} 1 & -2 & 1 \\ -2 & 5 & -4 \\ 1 & -4 & 6 \end{bmatrix}.$$

2.21. Exercícios variados sobre matrizes.

- Se uma matriz quadrada tem uma fila (linha ou coluna) de zeros, provar que é singular.
- Para cada uma das proposições seguintes relativas a matrizes $n \times n$, dar uma demonstração ou apresentar um contra exemplo.
 - Se $AB + BA = O$, então $A^2B^3 = B^3A^2$.
 - Se A e B são não singulares, então $A + B$ é não singular.
 - Se A e B são não singulares, então AB é não singular.
 - Se A , B e $A + B$ são não singulares, então $A - B$ é não singular.
 - Se $A^3 = O$, então $A - I$ é não singular.
 - Se o produto de k matrizes A_1, \dots, A_k é não singular, então cada matriz A_i é não singular.

$$3. \text{ Se } A = \begin{bmatrix} 1 & 2 \\ 5 & 4 \end{bmatrix} \text{ determinar uma matriz não singular } P \text{ tal que } P^{-1}AP = \begin{bmatrix} 6 & 0 \\ 0 & -1 \end{bmatrix}.$$

$$4. \text{ A matriz } A = \begin{bmatrix} a & i \\ i & b \end{bmatrix}, \text{ onde } i^2 = -1, a = \frac{1}{2}(1 + \sqrt{5}), \text{ e } b = \frac{1}{2}(1 - \sqrt{5}), \text{ goza da propriedade}$$

$A^2 = A$. Descrever completamente todas as matrizes 2×2 , A , com elementos complexos tais que $A^2 = A$.

$$5. \text{ Se } A^2 = A, \text{ provar que } (A + I)^k = I + (2^k - 1)A.$$

6. A teoria da relatividade restrita utiliza o conjunto de equações $x' = a(x - vt)$, $y' = y$, $z' = z$, $t' = a(t - vx/c^2)$, em que v representa a velocidade de uma partícula em movimento, c a velocidade da luz, e $a = c/\sqrt{c^2 - v^2}$, onde $|v| < c$. A transformação linear que aplica o vector bidimensional (x, t) em (x', t') chama-se uma *transformação de Lorentz*. A matriz correspondente relativa às bases usuais representa-se por $L(v)$ e é dada por

$$L(v) = a \begin{bmatrix} 1 & -v \\ -vc^{-2} & 1 \end{bmatrix}.$$

Observe-se que $L(v)$ é, não singular e que $L(0) = I$. Provar que $L(v)L(u) = L(w)$, onde $w = (u + v)c^2/(uv + c^2)$. Por outras palavras, o produto de duas transformações de Lorentz é outra transformação de Lorentz.

7. Se permutamos entre si as linhas e as colunas de uma matriz rectangular A , a nova matriz assim obtida diz-se a *transposta* de A e representa-se por A^t . Por exemplo, se tivermos

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad \text{então } A^t = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}.$$

Provar que a transposta goza das seguintes propriedades:

- (a) $(A^t)^t = A$. (b) $(A + B)^t = A^t + B^t$. (c) $(cA)^t = cA^t$.
 (d) $(AB)^t = B^t A^t$. (e) $(A^t)^{-1} = (A^{-1})^t$ se A é não singular.
8. Uma matriz quadrada A chama-se uma matriz *ortogonal* se $AA^t = I$. Verificar que a matriz 2×2 , $\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$, é ortogonal para cada real θ . Se A qualquer matriz ortogonal $n \times n$, provar que as suas linhas, consideradas como vectores de V_n , formam um conjunto ortogonal.
9. Para cada uma das seguintes proposições relativas a matrizes $n \times n$, dar uma demonstração ou então apresentar um contra exemplo.
 (a) Se A e B são ortogonais, então $A + B$ é ortogonal.
 (b) Se A e B são ortogonais, então AB é ortogonal.
 (c) Se A e B são ortogonais, então B é ortogonal.
10. Matrizes de Hadamard, assim designadas por Hadamard (1865-1963), são as matrizes $n \times n$ gozando das propriedades seguintes:
 I. Todo o elemento é 1 ou -1 .
 II. Cada linha, considerada como um vector de V_n , tem norma igual a \sqrt{n} .
 III. O produto escalar de duas quaisquer linhas distintas é 0. As matrizes de Hadamard aparecem em certos problemas de geometria e na teoria dos números, e foram recentemente aplicadas à construção da codificação óptica para a comunicação espacial. Apesar da sua aparente simplicidade, apresentam muitos problemas não resolvidos. O principal problema não resolvido, neste momento, consiste na determinação de todos os valores de n para os quais existe uma matriz de Hadamard, $n \times n$. Este exercício esboça uma resolução parcial.
 (a) Determinar todas as matrizes de Hadamard 2×2 (são precisamente 8).
 (b) Esta parte do exercício esboça uma demonstração simples do seguinte teorema: Se A é uma matriz de Hadamard $n \times n$, com $n > 2$, então n é múltiplo de 4. A demonstração baseia-se em dois lemas muito simples referentes ao espaço n dimensional. Demonstrar cada um desses lemas e aplicá-los às linhas da matriz de Hadamard para demonstrar o teorema.

LEMA 1. Se X, Y, Z são vectores ortogonais de V_n então tem-se

$$(X + Y) \cdot (X + Z) = \|X\|^2.$$

LEMA 2. Escrever $X = (x_1, \dots, x_n)$, $Y = (y_1, \dots, y_n)$, $Z = (z_1, \dots, z_n)$. Se cada componente x_i, y_i, z_i é ou 1 ou -1 , então o produto $(x_i + y_i)(x_i + z_i)$ é ou 0 ou 4.

DETERMINANTES

3.1. Introdução

Em muitas aplicações da álgebra linear à geometria e à análise, o conceito de determinante desempenha um papel importante. Neste capítulo estudam-se as propriedades fundamentais dos determinantes e algumas das suas aplicações.

Os determinantes de segunda e terceira ordem foram introduzidos no volume I como pretexto para uma notação conveniente, e sobretudo compacta, em certas fórmulas. Recordamos que um determinante de segunda ordem foi definido pela fórmula

$$(3.1) \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

A despeito da semelhança de notação, o determinante $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$ (escrito com traços verticais) é conceitualmente distinto da matriz $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ (escrita com parêntesis rectos). O determinante é um *número* atribuído à matriz e que se calcula segundo a fórmula (3.1). Para acentuar esta ligação também se escreve

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = \det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

No volume I definiram-se ainda os determinantes de terceira ordem em função de determinantes de segunda ordem pela fórmula

$$(3.2) \quad \det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}.$$

Este capítulo trata do problema mais geral, o determinante de uma matriz quadrada de ordem n com n um inteiro qualquer ≥ 1 . O nosso ponto de vista consiste em tratar o determinante como uma função que atribui a cada matriz quadrada A um número chamado o determinante de A e que se representa por $\det A$. É possível definir esta função por intermédio de uma fórmula que generaliza (3.1) e (3.2). Essa fórmula é uma soma de $n!$ produtos de elementos de A . Para grandes valores de n a fórmula é de difícil manejo e raramente se usa na prática. Parece assim preferível estudar determinantes de um outro ponto de vista, o qual porá em evidência as suas propriedades fundamentais. Estas propriedades, de grande importância nas aplicações, serão consideradas como *axiomas* para a função determinante. Em princípio, o nosso programa consiste de três pontos: (1) justificar a escolha dos axiomas; (2) deduzir outras propriedades dos determinantes a partir dos axiomas; (3) provar que existe uma e uma só função que satisfaz aos axiomas.

3.2. Justificação da escolha dos axiomas para a função determinante

No volume I provámos que o produto misto de três vectores A_1, A_2, A_3 no espaço E_3 pode exprimir-se como o determinante de uma matriz cujas linhas são as componentes dos vectores. Assim temos

$$A_1 \times A_2 \cdot A_3 = \det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix},$$

onde $A_1 = (a_{11}, a_{12}, a_{13})$, $A_2 = (a_{21}, a_{22}, a_{23})$, e $A_3 = (a_{31}, a_{32}, a_{33})$.

Se as linhas são linearmente independentes, o produto misto é diferente de zero; o valor absoluto do produto é igual ao volume do paralelepípedo definido pelos três vectores A_1, A_2, A_3 . Se as linhas são dependentes o produto misto é zero. Neste caso os vectores A_1, A_2 e A_3 são coplanares e o paralelepípedo degenera numa figura plana, com volume nulo.

Algumas das propriedades do produto misto servirão de justificação à escolha dos axiomas para a função determinante no caso de maior número de dimensões. Para estabelecer estas propriedades de uma maneira adequada à generalização, consideramos o produto misto como uma função dos três vectores linha A_1, A_2, A_3 . Representamos tal função por d ; assim,

$$d(A_1, A_2, A_3) = A_1 \times A_2 \cdot A_3.$$

Centramos a nossa atenção nas seguintes propriedades:

(a) *Homogeneidade em cada linha.* Por exemplo, a propriedade de homogeneidade na primeira linha significa que

$$d(tA_1, A_2, A_3) = t d(A_1, A_2, A_3) \quad \text{para todo escalar } t.$$

(b) *Aditividade em cada linha.* Por exemplo, a propriedade de aditividade na segunda linha significará que

$$d(A_1, A_2 + C, A_3) = d(A_1, A_2, A_3) + d(A_1, C, A_3)$$

para todo o vector C .

(c) *O produto misto é nulo se duas das filas são iguais.*

(d) *Normalização:*

$$d(i, j, k) = 1, \quad \text{onde } i = (1, 0, 0), \quad j = (0, 1, 0), \quad k = (0, 0, 1).$$

Cada uma destas propriedades pode ser facilmente verificada a partir das propriedades do produto escalar e produto vectorial. Algumas são sugeridas pela interpretação geométrica do produto misto como o volume do paralelepípedo determinado pelos vectores geométricos A_1, A_2, A_3 . O significado geométrico da propriedade aditiva (b) num caso particular é de especial interesse. Se tomarmos $C = A_1$ em (b), o segundo termo do segundo membro é nulo devido a (c), e a relação (b) vem

$$(3.3) \quad d(A_1, A_2 + A_1, A_3) = d(A_1, A_2, A_3).$$

Esta propriedade está representada na figura 3.1. na qual se mostra um paralelepípedo definido por A_1, A_2, A_3 , e outro paralelepípedo definido por $A_1, A_1 + A_2, A_3$. A equação (3.3) estabelece, apenas, que estes dois paralelepípedos têm iguais volumes. Isto, geometricamente, é evidente devido ao facto de que os paralelepípedos têm bases de áreas iguais e a mesma altura.

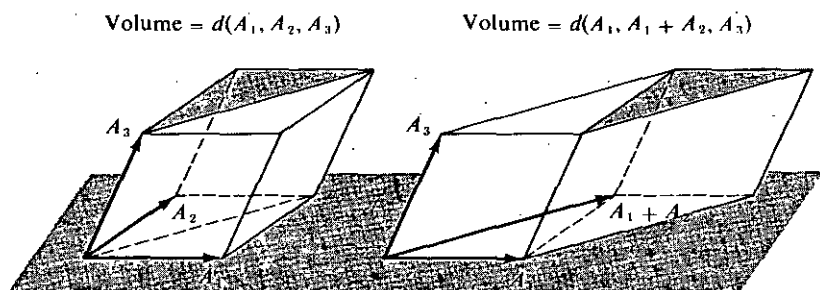


FIG. 3.1. Interpretação geométrica da propriedade $d(A_1, A_2, A_3) = d(A_1, A_1 + A_2, A_3)$.
Os dois paralelepípedos têm volumes iguais.

3.3. Um conjunto de axiomas para a função determinante

As propriedades do produto misto mencionadas na Secção anterior podem ser convenientemente generalizadas e usadas como axiomas para determinantes de ordem n . Se $A = (a_{ij})$ é uma matriz $n \times n$ com elementos reais ou complexos, representemos as suas linhas por A_1, \dots, A_n . Assim, a linha de ordem i de A é um vector no espaço E_n dimensional definido por

$$A_i = (a_{i1}, a_{i2}, \dots, a_{in}).$$

Consideremos o determinante como uma função das n linhas A_1, \dots, A_n e representemos o seu valor por $d(A_1, \dots, A_n)$ ou por $\det A$.

DEFINIÇÃO AXIOMÁTICA DE UMA FUNÇÃO DETERMINANTE. *Uma função d de valores reais ou complexos, definida para cada sistema ordenado de n vectores A_1, \dots, A_n de E_n , diz-se uma função determinante de ordem n se verifica os seguintes axiomas, quaisquer que sejam os vectores A_1, \dots, A_n e C em E_n :*

AXIOMA 1. HOMOGENEIDADE EM CADA LINHA. *Se a linha de ordem k , A_k , se multiplica por um escalar t , então o determinante vem também multiplicado por t :*

$$d(\dots, tA_k, \dots) = t d(\dots, A_k, \dots).$$

AXIOMA 2. ADITIVIDADE EM CADA LINHA. *Para todo k tem-se*

$$d(A_1, \dots, A_k + C, \dots, A_n) = d(A_1, \dots, A_k, \dots, A_n) + d(A_1, \dots, C, \dots, A_n).$$

AXIOMA 3. O DETERMINANTE ANULA-SE SE DUAS LINHAS SÃO IGUAIS:

$$d(A_1, \dots, A_n) = 0 \quad \text{se } A_i = A_j \text{ para quaisquer } i \text{ e } j, \text{ com } i \neq j.$$

AXIOMA 4. O DETERMINANTE DA MATRIZ IDENTIDADE É IGUAL A 1:

$$d(I_1, \dots, I_n) = 1, \text{ com } I_k \text{ o vector coordenado unitário de ordem } k.$$

Os dois primeiros axiomas afirmam que o determinante de uma matriz é uma função linear de cada uma das linhas. Isto é muitas vezes referido dizendo que o determinante é uma função *multilinear* das suas linhas. Por aplicação reiterada da propriedade de linearidade, a primeira linha pode escrever-se

$$d\left(\sum_{k=1}^p t_k C_k, A_2, \dots, A_n\right) = \sum_{k=1}^p t_k d(C_k, A_2, \dots, A_n),$$

onde t_1, \dots, t_p são escalares e C_1, \dots, C_p são vectores quaisquer de E_n

Por vezes utiliza-se uma forma mais fraca do Axioma 3:

AXIOMA 3'. O DETERMINANTE ANULA-SE SE DUAS LINHAS CONSECUTIVAS SÃO IGUAIS.

$$d(A_1, \dots, A_n) = 0 \quad \text{se} \quad A_k = A_{k+1} \quad \text{para algum } k = 1, 2, \dots, n-1.$$

É um facto notável que para um dado n existe uma e uma só função d que verifique os axiomas 1, 2, 3' e 4. A prova deste facto, um dos resultados importantes deste capítulo, será apresentada adiante. O teorema seguinte dá-nos propriedades dos determinantes deduzidas unicamente a partir dos Axiomas 1, 2 e 3'. Uma destas propriedades é o Axioma 3. Deve ter-se presente que o Axioma 4 não intervém na demonstração do teorema. Esta observação será útil mais tarde quando se demonstrar a unicidade da função determinante.

TEOREMA 3.1. *Uma função determinante satisfazendo aos Axiomas 1, 2 e 3' possui ainda mais as seguintes propriedades:*

(a) *O determinante anula-se se tiver alguma linha de zeros:*

$$d(A_1, \dots, A_n) = 0 \quad \text{se} \quad A_k = 0 \quad \text{para quaisquer } k.$$

(b) *O determinante muda de sinal se permutam duas filas consecutivas:*

$$d(\dots, A_k, A_{k+1}, \dots) = -d(\dots, A_{k+1}, A_k, \dots).$$

(c) *O determinante muda de sinal se duas quaisquer linhas A_i e A_j , com $i \neq j$, se permutam.*

(d) *O determinante anula-se se duas quaisquer linhas são iguais:*

$$d(A_1, \dots, A_n) = 0 \quad \text{se} \quad A_i = A_j \quad \text{quaisquer que sejam } i \text{ e } j \text{ com } i \neq j.$$

(e) *O determinante anula-se se as suas linhas são linearmente dependentes.*

Demonstração. Para demonstrar (a) basta tomar $t = 0$ no Axioma 1. Para demonstrar (b), designamos por B a matriz com as linhas iguais às de A , excepto para as linhas de ordem k e $k+1$. Sejam as linhas B_k e B_{k+1} iguais a $A_k + A_{k+1}$. Então devido ao Axioma 3', $\det B = 0$. Podemos pois escrever

$$d(\dots, A_k + A_{k+1}, A_k + A_{k+1}, \dots) = 0.$$

Aplicando a propriedade aditiva às linhas k e $k + 1$ podemos de novo escrever a igualdade anterior na forma

$$d(\dots, A_k, A_k, \dots) + d(\dots, A_k, A_{k+1}, \dots) + d(\dots, A_{k+1}, A_k, \dots) + d(\dots, A_{k+1}, A_{k+1}, \dots) = 0$$

Os primeiros e quarto termos são nulos devido ao Axioma 3'. Por conseguinte os segundo e terceiro termos são simétricos, o que demonstra (b).

Para provar (c) podemos admitir que $i < j$. Podemos permutar as linhas A_i e A_j efectuando um número ímpar de permutações de linhas consecutivas. Permutamos em primeiro lugar A_j sucessivamente com cada uma das linhas que a precedem $A_{j-1}, A_{j-2}, \dots, A_i$. No total temos $j - i$ permutações. Em seguida permutamos A_i sucessivamente com as linhas que se seguem $A_{i+1}, A_{i+2}, \dots, A_{j-1}$. No total temos $j - i - 1$ permutações de linhas consecutivas. Em cada uma de três permutações o sinal do determinante muda. Uma vez que são $(j - i) + (j - i - 1) = 2(j - i) - 1$ permutações ao todo (um número ímpar), o determinante muda de sinal um número ímpar de vezes, o que demonstra (c).

Para demonstrar (d), seja B a matriz deduzida de A por permutação das linhas A_i e A_j . Uma vez que $A_i = A_j$ tem-se $B = A$ e por conseguinte $\det B = \det A$. Mas por (c), $\det B = -\det A$ e consequentemente $\det A = 0$.

Para a demonstração de (e) admite-se a existência de escalares c_1, c_2, \dots, c_n , não todos nulos, tais que $\sum_{k=1}^n c_k A_k = O$. Então qualquer linha A_k com $c_k \neq 0$ pode exprimir-se como uma combinação linear das outras linhas. Por comodidade, admitamos que A_1 é uma combinação linear das restantes, por exemplo $A_1 = \sum_{k=2}^n t_k A_k$. Devido à propriedade de linearidade da primeira linha temos

$$d(A_1, A_2, \dots, A_n) = d\left(\sum_{k=2}^n t_k A_k, A_2, \dots, A_n\right) = \sum_{k=2}^n t_k d(A_k, A_2, \dots, A_n).$$

Mas cada termo $d(A_k, A_2, \dots, A_n)$ na última soma é zero, visto ser A_k igual a pelo menos um dos A_2, \dots, A_n . Por tal razão a soma total é zero. Se a linha A_i é uma combinação linear das outras linhas, raciocinamos da mesma maneira, utilizando a linearidade na linha de ordem i , e assim está demonstrada (e).

3.4. Cálculo de determinantes

Ao chegarmos a este ponto será conveniente calcularmos alguns determinantes, servindo-nos unicamente dos axiomas e das propriedades dadas no teorema 3.1, presumindo sempre que as funções determinante existem. Em cada um dos exemplos apresentados não aplicaremos o Axioma 4 até à derradeira fase do Cálculo.

EXEMPLO 1. *Determinante de uma matriz 2×2 . Vamos provar que*

$$(3.4) \quad \det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

Escrevamos os vectores linha como combinações lineares dos vectores unitários coordenados $i = (1, 0)$ e $j = (0, 1)$:

$$A_1 = (a_{11}, a_{12}) = a_{11}i + a_{12}j, \quad A_2 = (a_{21}, a_{22}) = a_{21}i + a_{22}j.$$

Usando a linearidade na primeira linha temos

$$d(A_1, A_2) = d(a_{11}i + a_{12}j, A_2) = a_{11}d(i, A_2) + a_{12}d(j, A_2).$$

Usando agora a linearidade na segunda linha obtemos

$$d(i, A_2) = d(i, a_{21}i + a_{22}j) = a_{21}d(i, i) + a_{22}d(i, j) = a_{22}d(i, j),$$

Do mesmo modo se encontra

$$d(j, A_2) = d(j, a_{21}i + a_{22}j) = a_{21}d(j, i) = -a_{21}d(i, j).$$

Finalmente, obtemos

$$d(A_1, A_2) = (a_{11}a_{22} - a_{12}a_{21})d(i, j).$$

Mas $d(i, j) = 1$ pelo Axioma 4, pelo que $d(A_1, A_2) = a_{11}a_{22} - a_{12}a_{21}$, como se afirmara.

Este raciocínio mostra que se existe uma função determinante para matrizes 2×2 , então ela terá necessariamente a forma (3.4). Reciprocamente, é fácil verificar que esta fórmula define, na realidade, uma função determinante de ordem 2. Por conseguinte está demonstrado que existe uma e uma só fracção determinante de ordem 2.

EXEMPLO 2. *Determinante de uma matriz diagonal.* Uma matriz quadrada da forma

$$A = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}$$

diz-se uma *matriz diagonal*. Cada elemento a_{ij} não pertencendo à diagonal ($i \neq j$) é zero. Provaremos que o determinante de A é igual ao produto dos elementos pertencentes à *diagonal principal* (ou *primeira diagonal*),

$$(3.5) \quad \det A = a_{11}a_{22} \cdots a_{nn}.$$

A linha de ordem k de A é o produto de um escalar pelo k -ésimo vector coordenado unitário, $A_k = a_{kk}I_k$. Aplicando a propriedade da homogeneidade, repetidas vezes, obtemos

$$\det A = d(A_1, \dots, A_n) = d(a_{11}I_1, \dots, a_{nn}I_n) = a_{11} \cdots a_{nn} d(I_1, \dots, I_n).$$

Esta fórmula pode escrever-se

$$\det A = a_{11} \cdots a_{nn} \det I,$$

com I a matriz identidade. O Axioma 4, porém, diz-nos que $\det I = 1$, pelo que se obtém (3.5).

EXEMPLO 3. *Determinante de uma matriz triangular superior.* Uma matriz quadrada da forma

$$U = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & u_{nn} \end{bmatrix}$$

diz-se uma matriz triangular superior. Todos os elementos abaixo da diagonal principal são nulos. Vamos demonstrar que o determinante de tal matriz é igual ao produto dos elementos da sua diagonal principal.

$$\det U = u_{11}u_{22} \cdots u_{nn}.$$

Provemos em primeiro lugar que $\det U = 0$ se algum elemento da diagonal $u_{ii} = 0$. Se fôr nulo o último elemento u_{nn} , então a última linha é O e $\det U = 0$, devido ao Teorema 3.1 (a). Suponhamos, então, que algum dos elementos precedentes na diagonal, u_{ii} , é zero. Para especificar, admitamos $u_{22} = 0$. Então cada um dos $n - 1$ vectores linhas U_2, \dots, U_n tem as duas primeiras componentes nulas. Daqui resulta que estes vectores geram um subespaço de dimensão no máximo, $n - 2$. Deste modo estas $n - 1$ linhas (e consequentemente *todas* as linhas) são dependentes. Pelo teorema 3.1 (e), $\det U = 0$. Do mesmo modo verificamos que $\det U = 0$ se for nulo qualquer elemento da diagonal principal.

Consideremos agora o caso geral. Seja a primeira linha U_1 definida como uma soma de dois vectores-linha,

$$U_1 = V_1 + V'_1,$$

com $V_1 = [u_{11}, 0, \dots, 0]$ e $V'_1 = [0, u_{12}, u_{13}, \dots, u_{1n}]$. Devido à linearidade na primeira linha temos

$$\det U = \det (V_1, U_2, \dots, U_n) + \det (V'_1, U_2, \dots, U_n).$$

Mas $\det(V'_1, U_2, \dots, U_n) = 0$ visto que este é o determinante de uma matriz triangular superior com um elemento da diagonal principal igual a zero. Consequentemente, temos

$$(3.6) \quad \det U = \det(V_1, U_2, \dots, U_n).$$

Tratamos agora o vector-linha U_2 de uma maneira análoga, exprimindo-o como uma soma,

$$U_2 = V_2 + V'_2,$$

onde

$$V_2 = [0, u_{22}, 0, \dots, 0] \quad \text{e} \quad V'_2 = [0, 0, u_{23}, \dots, u_{2n}].$$

Substituindo no segundo membro de (3.6) e aplicando a propriedade da linearidade na segunda linha obtemos

$$(3.7) \quad \det U = \det(V_1, V_2, U_3, \dots, U_n),$$

visto que $\det(V_1, V'_2, U_3, \dots, U_n) = 0$. Repetindo o raciocínio para cada uma das linhas seguintes no segundo membro de (3.7) obtemos finalmente

$$\det U = \det(V_1, V_2, \dots, V_n),$$

com (V_1, V_2, \dots, V_n) uma matriz diagonal com os mesmos elementos que U na diagonal principal. Por conseguinte, segundo o exemplo 2, temos

$$\det U = u_{11}u_{22} \cdots u_{nn},$$

como pretendíamos demonstrar.

EXEMPLO 4. Cálculo pelo método de Gauss-Jordan. O método de eliminação de Gauss-Jordan para resolver sistemas de equações lineares é igualmente um dos melhores métodos para o cálculo de determinantes. Recorde-se que o método consiste na aplicação de três tipos de operações às linhas da matriz:

- (1) *Permutação de duas linhas.*
- (2) *Multiplicação de todos os elementos de uma linha por um escalar não nulo.*
- (3) *Adição a outra de uma linha multiplicada por um escalar.*

Efectuando estas operações de uma maneira sistemática podemos transformar qualquer matriz quadrada A numa matriz triangular superior U cujo determinante já sabemos calcular. É fácil determinar a relação entre $\det A$ e $\det U$. De cada vez que se efectua a operação (1) o determinante muda de sinal. De cada vez que se efectua (2), sendo c um escalar não nulo, $c \neq 0$, o determinante vem multiplicado por c . De cada vez que se efectua (3) o determinante permanece inalterado. Desta maneira, se efectuarmos p

vezes a operação (1) e se c_1, \dots, c_q são escalares diferentes de zero usados em q operações do tipo (2), temos

$$(3.8) \quad \det A = (-1)^p (c_1 c_2 \cdots c_q)^{-1} \det U.$$

Mais uma vez observamos que esta fórmula é uma consequência unicamente dos três primeiros axiomas. A sua demonstração não fez intervir o Axioma 4.

3.5. O teorema de unicidade

No Exemplo 3 da secção anterior provou-se que os Axiomas 1, 2 e 3 implicam a fórmula $\det U = u_{11}u_{22}\cdots u_{nn} \det I$. Combinando-a com (3.8) vemos que para toda a matriz A , $n \times n$, existe um escalar c (que depende de A) tal que

$$(3.9) \quad d(A_1, \dots, A_n) = c d(I_1, \dots, I_n).$$

Além disso, esta fórmula é uma consequência dos Axiomas 1, 2 e 3 unicamente. A partir deste resultado podemos facilmente provar que não existe mais do que uma função determinante.

TEOREMA 3.2. TEOREMA DE UNICIDADE PARA DETERMINANTES. *Se d é uma função satisfazendo aos quatro axiomas para uma função determinante de ordem n , e f é outra função satisfazendo apenas aos Axiomas 1, 2 e 3, então para qualquer escolha de um sistema de vectores A_1, \dots, A_n no espaço n dimensional tem-se*

$$(3.10) \quad f(A_1, \dots, A_n) = d(A_1, \dots, A_n) f(I_1, \dots, I_n).$$

Em particular, se f também verifica o Axioma 4 então será

$$f(A_1, \dots, A_n) = d(A_1, \dots, A_n).$$

Demonstração. Seja $g(A_1, \dots, A_n) = f(A_1, \dots, A_n) - d(A_1, \dots, A_n) f(I_1, \dots, I_n)$. Vamos provar que $g(A_1, \dots, A_n) = 0$ qualquer que seja a escolha de A_1, \dots, A_n . Uma vez que, quer d , quer f , verificam os axiomas 1, 2 e 3, o mesmo é verdadeiro para g . Por conseguinte g satisfaz também à equação (3.9) que fora deduzida unicamente a partir dos primeiros três axiomas. Deste modo podemos escrever

$$(3.11) \quad g(A_1, \dots, A_n) = c g(I_1, \dots, I_n),$$

com c um escalar dependendo de A . Tomando $A = I$ na definição de g e observando que d satisfaz ao Axioma 4 encontramos

$$g(I_1, \dots, I_n) = f(I_1, \dots, I_n) - f(I_1, \dots, I_n) = 0.$$

Por conseguinte, a equação (3.11) vem $g(A_1, A_2, \dots, A_n) = 0$, e o teorema está demonstrado.

3.6. Exercícios

Neste conjunto de exercícios deve admitir-se a existência da função determinante. Os determinantes de terceira ordem podem ser calculados pela fórmula (3.2).

1. Calcular cada um dos determinantes seguintes:

$$(a) \begin{vmatrix} 2 & 1 & 1 \\ 1 & 4 & -4 \\ 1 & 0 & 2 \end{vmatrix}, \quad (b) \begin{vmatrix} 3 & 0 & 8 \\ 5 & 0 & 7 \\ -1 & 4 & 2 \end{vmatrix}, \quad (c) \begin{vmatrix} a & 1 & 0 \\ 2 & a & 2 \\ 0 & 1 & a \end{vmatrix}.$$

2. Se $\det \begin{bmatrix} x & y & z \\ 3 & 0 & 2 \\ 1 & 1 & 1 \end{bmatrix} = 1$, calcular o determinante de cada uma das matrizes seguintes:

$$(a) \begin{bmatrix} 2x & 2y & 2z \\ \frac{3}{2} & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad (b) \begin{bmatrix} x & y & z \\ 3x+3 & 3y & 3z+2 \\ x+1 & y+1 & z+1 \end{bmatrix}, \quad (c) \begin{bmatrix} x-1 & y-1 & z-1 \\ 4 & 1 & 3 \\ 1 & 1 & 1 \end{bmatrix}.$$

3. (a) Provar que $\begin{vmatrix} 1 & 1 & 1 \\ a & b & c \\ a^2 & b^2 & c^2 \end{vmatrix} = (b-a)(c-a)(c-b)$.

(b) Determinar as fórmulas correspondentes para as determinantes

$$\begin{vmatrix} 1 & 1 & 1 \\ a & b & c \\ a^3 & b^3 & c^3 \end{vmatrix} \quad e \quad \begin{vmatrix} 1 & 1 & 1 \\ a^2 & b^2 & c^2 \\ a^3 & b^3 & c^3 \end{vmatrix}.$$

4. Calcular o determinante de cada uma das matrizes seguintes, pela transformação de cada uma delas numa matriz triangular superior

$$(a) \begin{bmatrix} 1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 \end{bmatrix}, \quad (b) \begin{bmatrix} 1 & 1 & 1 & 1 \\ a & b & c & d \\ a^2 & b^2 & c^2 & d^2 \\ a^3 & b^3 & c^3 & d^3 \end{bmatrix}, \quad (c) \begin{bmatrix} 1 & 1 & 1 & 1 \\ a & b & c & d \\ a^2 & b^2 & c^2 & d^2 \\ a^4 & b^4 & c^4 & d^4 \end{bmatrix}.$$

$$(d) \begin{bmatrix} a & 1 & 0 & 0 & 0 \\ 4 & a & 2 & 0 & 0 \\ 0 & 3 & a & 3 & 0 \\ 0 & 0 & 2 & a & 4 \\ 0 & 0 & 0 & 1 & a \end{bmatrix}, \quad (e) \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 \end{bmatrix}$$

5. Uma matriz triangular inferior $A = (a_{ij})$ é uma matriz quadrada em que os elementos acima da diagonal principal são todos nulos, isto é, $a_{ij} = 0$ sempre que $i < j$. Provar que determinante de tal matriz é igual ao produto dos respectivos elementos da diagonal principal: $\det A = a_{11}a_{22} \cdots a_{nn}$.
6. Sejam f_1, f_2, g_1, g_2 quatro funções deriváveis no intervalo (a, b) . Defina-se

$$F(x) = \begin{vmatrix} f_1(x) & f_2(x) \\ g_1(x) & g_2(x) \end{vmatrix}$$

para todo x de (a, b) . Provar que

$$F'(x) = \begin{vmatrix} f_1'(x) & f_2'(x) \\ g_1(x) & g_2(x) \end{vmatrix} + \begin{vmatrix} f_1(x) & f_2(x) \\ g_1'(x) & g_2'(x) \end{vmatrix}.$$

7. Estabelecer e demonstrar uma generalização do Exercício 6 para o determinante

$$F(x) = \begin{vmatrix} f_1(x) & f_2(x) & f_3(x) \\ g_1(x) & g_2(x) & g_3(x) \\ h_1(x) & h_2(x) & h_3(x) \end{vmatrix}.$$

8. (a) Se $F(x) = \begin{vmatrix} f_1(x) & f_2(x) \\ f_1'(x) & f_2'(x) \end{vmatrix}$, provar que $F'(x) = \begin{vmatrix} f_1(x) & f_2(x) \\ f_1''(x) & f_2''(x) \end{vmatrix}$.
- (b) Estabelecer e demonstrar o resultado correspondente para determinantes de terceira ordem, admitindo a validade da equação (3.2).
9. Sejam U e V duas matrizes triangulares superiores $n \times n$.
- (a) Provar que cada $U + V$ e UV é uma matriz triangular superior.
- (b) Provar que $\det(UV) = (\det U)(\det V)$.
- (c) Se $\det U \neq 0$, provar que existe uma matriz triangular superior U^{-1} tal que $UU^{-1} = I$, e demonstrar que $\det(U^{-1}) = 1/\det U$.
10. Calcular $\det A$, $\det(A^{-1})$ e A^{-1} para a seguinte matriz triangular superior:

$$A = \begin{bmatrix} 2 & 3 & 4 & 5 \\ 0 & 2 & 3 & 4 \\ 0 & 0 & 2 & 3 \\ 0 & 0 & 0 & 2 \end{bmatrix}.$$

3.7. Produto de determinantes

Nesta Secção usamos o teorema da unicidade para demonstrar que o determinante de um produto de duas matrizes quadradas é igual ao produto dos correspondentes determinantes dessas matrizes,

$$\det(AB) = (\det A)(\det B),$$

pressupondo evidentemente que a função determinante existe.

Recordamos que o produto AB de duas matrizes $A = (a_{ij})$ e $B = (b_{ij})$ é a matriz $C = (c_{ij})$ cujo elemento c_{ij} é dado por

$$(3.12) \quad c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}.$$

O produto de matrizes define-se unicamente se o número de colunas do factor da esquerda, A , é igual ao número de linhas do factor da direita, B . Isto verifica-se sempre que A e B são matrizes quadradas da mesma ordem.

A demonstração da fórmula para o produto fará uso de uma relação simples definida entre as linhas de AB e as de A . Vamos estabelecer-las como um lema. Como habitualmente, representamos por A_i a linha de ordem i da matriz A .

LEMA 3.3. *Se A é uma matriz $m \times n$ e B outra matriz $n \times p$, então tem-se*

$$(AB)_i = A_i B.$$

isto é, a linha de ordem i do produto AB é igual ao produto da linha A_i de A pela matriz B .

Demonstração. Represente B^j a coluna de ordem j de B e seja $C = AB$. Então a soma (3.12) pode considerar-se como o produto escalar da linha de ordem i de A pela coluna de ordem j de B ,

$$c_{ij} = A_i \cdot B^j.$$

Portanto a linha de ordem i de C , é a matriz linha

$$C_i = [A_i \cdot B^1, A_i \cdot B^2, \dots, A_i \cdot B^p].$$

Mas este é também o resultado da multiplicação da matriz linha A_i por B , visto que

$$A_i B = [a_{i1}, \dots, a_{in}] \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1p} \\ \vdots & \vdots & & \vdots \\ b_{n1} & b_{n2} & \dots & b_{np} \end{bmatrix} = [A_i \cdot B^1, \dots, A_i \cdot B^p].$$

Consequentemente $C_i = A_i B$ e o lema está demonstrado.

TEOREMA 3.4. FÓRMULA PARA O PRODUTO DE DETERMINANTES. *Para duas quaisquer matrizes $n \times n$, A e B , tem-se*

$$\det(AB) = (\det A)(\det B).$$

Demonstração. Uma vez que $(AB)_i = A_i B$, temos que demonstrar que

$$d(A_1 B, \dots, A_n B) = d(A_1, \dots, A_n) d(B_1, \dots, B_n).$$

Recorrendo de novo ao lema, temos $B_i = (IB)_i = I_i B$, com I a matriz identidade $n \times n$. Portanto $d(B_1, \dots, B_n) = d(I_1 B, \dots, I_n B)$, e temos que provar que

$$d(A_1 B, \dots, A_n B) = d(A_1, \dots, A_n) d(I_1 B, \dots, I_n B).$$

Consideremos B fixo e introduzamos uma função f definida pela fórmula

$$f(A_1, \dots, A_n) = d(A_1 B, \dots, A_n B).$$

A equação que pretendemos demonstrar estabelece que

$$(3.13) \quad f(A_1, \dots, A_n) = d(A_1, \dots, A_n) f(I_1, \dots, I_n).$$

E é agora uma questão simples comprovar que f satisfaz aos Axiomas 1, 2 e 3 relativos à função determinante pelo que, devido ao teorema da unicidade, a igualdade (3.13) é válida para uma matriz qualquer A , ficando assim completada a demonstração.

Apresentamos a seguir aplicações da fórmula do produto.

3.8. Determinante da matriz inversa de uma matriz não singular

Recordemos que uma matriz quadrada A diz-se não singular se admite matriz inversa esquerda B tal que $BA = I$. Se A admite uma inversa esquerda ela é única e é também inversa direita, $AB = I$. Representemos a inversa por A^{-1} . A relação entre $\det A$ e $\det A^{-1}$ é a mais natural que se podia esperar.

TEOREMA 3.5. *Se a matriz A é não singular, então $\det A \neq 0$ e tem-se*

$$(3.14) \quad \det A^{-1} = \frac{1}{\det A}.$$

Demonstração. A partir da fórmula do produto temos

$$(\det A)(\det A^{-1}) = \det (AA^{-1}) = \det I = 1.$$

Por conseguinte $\det A \neq 0$ e (3.14) é válida.

O teorema 3.5 mostra que o não anulamento de $\det A$ é uma condição necessária para que A seja não singular. Provaremos mais adiante que esta condição é também suficiente, isto é, se $\det A \neq 0$ então A^{-1} existe.

3.9. Determinantes e independência de vectores

Do teorema 3.5 pode deduzir-se um critério simples para averiguar a independência de vectores.

TEOREMA 3.6. *Um conjunto de n vectores A_1, \dots, A_n de um espaço n -dimensional é linearmente independente se e só se $d(A_1, \dots, A_n) \neq 0$.*

Demonstração. Já provámos no teorema 3.2 (e) que a dependência implica $d(A_1, \dots, A_n) = 0$. Para provar a inversa, admitamos que A_1, \dots, A_n são independentes e demonstremos que $d(A_1, \dots, A_n) \neq 0$.

Represente V_n o espaço linear dos n -tuplos de escalares. Uma vez que A_1, \dots, A_n são n -elementos independentes de um espaço n -dimensional, formam uma base para V_n . Pelo teorema 2.12 existe uma transformação linear $T: V_n \rightarrow V_n$, a qual aplica estes n vectores nos vectores coordenados unitários,

$$T(A_k) = I_k \quad \text{para } k = 1, 2, \dots, n.$$

Por conseguinte existe uma matriz $n \times n$, B , tal que

$$A_k B = I_k \quad \text{para } k = 1, 2, \dots, n.$$

Mas pelo lema 3.3 temos $A_k B = (AB)_k$, com A a matriz cujas linhas são A_1, \dots, A_n . Consequentemente $AB = I$, pelo que A é não singular e $\det A \neq 0$.

3.10. Determinante de uma matriz diagonal por blocos

Uma matriz quadrada C da forma

$$C = \begin{bmatrix} A & O \\ O & B \end{bmatrix},$$

onde A e B são matrizes quadradas e cada O representa uma matriz de zeros diz-se uma *matriz diagonal por blocos*, com dois blocos diagonais A e B . Como exemplo considere-se a matriz 5×5

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 4 & 5 & 6 \\ 0 & 0 & 7 & 8 & 9 \end{bmatrix}.$$

Os blocos diagonais são neste caso

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{e} \quad B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}.$$

O teorema que se segue mostra que o determinante de uma matriz diagonal por blocos é igual ao produto dos determinantes dos respectivos blocos diagonais.

TEOREMA 3.7. *Quaisquer que sejam as matrizes quadradas A e B tem-se*

$$(3.15) \quad \det \begin{bmatrix} A & O \\ O & B \end{bmatrix} = (\det A)(\det B).$$

Demonstração. Admitamos que A é $n \times n$ e B é $m \times m$. Observe-se que a matriz diagonal por blocos pode exprimir-se como um produto da forma

$$\begin{bmatrix} A & O \\ O & B \end{bmatrix} = \begin{bmatrix} A & O \\ O & I_m \end{bmatrix} \begin{bmatrix} I_n & O \\ O & B \end{bmatrix}$$

onde I_n e I_m são matrizes identidade de ordens n e m respectivamente. Por conseguinte, segundo a fórmula do produto de determinantes temos

$$(3.16) \quad \det \begin{bmatrix} A & O \\ O & B \end{bmatrix} = \det \begin{bmatrix} A & O \\ O & I_m \end{bmatrix} \det \begin{bmatrix} I_n & O \\ O & B \end{bmatrix}.$$

Consideremos agora o determinante $\det \begin{bmatrix} A & O \\ O & I_m \end{bmatrix}$ como uma função das n linhas de

A . Tal é possível devido ao bloco de zeros do canto superior direito. É fácil verificar que esta função satisfaz aos quatro axiomas para uma função determinante de ordem n . Deste modo, pelo teorema da unicidade, devemos ter

$$\det \begin{bmatrix} A & O \\ O & I_m \end{bmatrix} = \det A.$$

Um argumento semelhante mostra que $\det \begin{bmatrix} I_n & O \\ O & B \end{bmatrix} = \det B$. Consequentemente (3.16) implica (3.15).

3.11. Exercícios

- Para cada uma das proposições seguintes, referentes a matrizes quadradas, efectuar a correspondente demonstração ou apresentar um contra exemplo.
 - $\det(A + B) = \det A + \det B$.
 - $\det\{(A + B)^2\} = \{\det(A + B)\}^2$
 - $\det\{(A + B)^2\} = \det(A^2 + 2AB + B^2)$
 - $\det\{(A + B)^2\} = \det(A^2 + B^2)$.
- (a) Estabelecer o teorema 3.7 para matrizes diagonais por blocos, com três blocos diagonais:

$$\det \begin{bmatrix} A & O & O \\ O & B & O \\ O & O & C \end{bmatrix} = (\det A)(\det B)(\det C).$$

- Generalizar e demonstrar o teorema para matrizes diagonais por blocos, com qualquer número de blocos diagonais.

$$3. \text{ Seja } A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ a & b & c & d \\ e & f & g & h \end{bmatrix}, \quad B = \begin{bmatrix} a & b & c & d \\ e & f & g & h \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \text{ Provar que } \det A = \det \begin{bmatrix} c & d \\ g & h \end{bmatrix} \text{ e que}$$

$$\det B = \det \begin{bmatrix} a & b \\ e & f \end{bmatrix}.$$

- Estabelecer e demonstrar uma generalização do Exercício 3 para matrizes $n \times n$.

$$5. \text{ Seja } A = \begin{bmatrix} a & b & 0 & 0 \\ c & d & 0 & 0 \\ e & f & g & h \\ x & y & z & w \end{bmatrix}. \text{ Provar que } \det A = \det \begin{bmatrix} a & b \\ c & d \end{bmatrix} \det \begin{bmatrix} g & h \\ z & w \end{bmatrix}.$$

- Estabelecer e demonstrar uma generalização do Exercício 5 para matrizes $n \times n$ da forma

$$A = \begin{bmatrix} B & O \\ C & D \end{bmatrix}$$

onde B, C, D representam matrizes quadradas e O representa uma matriz de zeros.

7. Servindo-se do teorema 3.6, verificar se os seguintes conjuntos de vectores são linearmente dependentes ou independentes.

(a) $A_1 = (1, -1, 0)$, $A_2 = (0, 1, -1)$, $A_3 = (2, 3, -1)$.

(b) $A_1 = (1, -1, 2, 1)$, $A_2 = (-1, 2, -1, 0)$, $A_3 = (3, -1, 1, 0)$, $A_4 = (1, 0, 0, 1)$.

(c) $A_1 = (1, 0, 0, 0, 1)$, $A_2 = (1, 1, 0, 0, 0)$, $A_3 = (1, 0, 1, 0, 1)$, $A_4 = (1, 1, 0, 1, 1)$, $A_5 = (0, 1, 0, 1, 0)$.

3.12. Fórmulas para o desenvolvimento de determinantes.

Menores e complementos algébricos

Ainda não demonstrámos que a função determinante existe, excepto para o caso de uma matriz 2×2 . Nesta Secção vamos explorar a propriedade de linearidade e o teorema da unicidade para demonstrarmos que, se os determinantes existem, eles podem ser calculados por intermédio de uma fórmula que exprime cada determinante de ordem n como uma combinação linear de determinantes de ordem $n-1$. A equação (3.2), da Secção 3.1, é um exemplo desta fórmula no caso 3×3 . A fórmula geral sugerirá um método para demonstrar a existência de funções determinantes por indução.

Cada linha de uma matriz A , $n \times n$, pode exprimir-se como uma combinação linear dos n vectores coordenados unitários I_1, \dots, I_n . Por exemplo, a primeira linha A_1 pode escrever-se

$$A_1 = \sum_{j=1}^n a_{1j} I_j.$$

Uma vez que os determinantes são lineares a respeito da primeira linha temos

$$(3.17) \quad d(A_1, A_2, \dots, A_n) = d\left(\sum_{j=1}^n a_{1j} I_j, A_2, \dots, A_n\right) = \sum_{j=1}^n a_{1j} d(I_j, A_2, \dots, A_n).$$

Quer dizer, para calcular $\det A$ basta calcular $d(I_j, A_2, \dots, A_n)$ para cada vector coordenado unitário I_j .

Utilizemos a notação A'_{1j} para representar a matriz obtida de A pela substituição da primeira linha A_1 pelo vector unitário I_j . Por exemplo, se $n = 3$ existem três matrizes desse tipo:

$$A'_{11} = \begin{bmatrix} 1 & 0 & 0 \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \quad A'_{12} = \begin{bmatrix} 0 & 1 & 0 \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \quad A'_{13} = \begin{bmatrix} 0 & 0 & 1 \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

Observe-se que $\det A'_{1j} = d(I_j, A_2, \dots, A_n)$. A equação (3.17) pode agora escrever-se na forma

$$(3.18) \quad \det A = \sum_{j=1}^n a_{1j} \det A'_{1j}.$$

Esta é uma *fórmula de desenvolvimento* do determinante; exprime o determinante de A como uma combinação linear dos elementos da sua primeira linha.

O raciocínio efectuado para deduzir (3.18) pode aplicar-se à linha de ordem k em vez da primeira linha. O resultado será uma fórmula de desenvolvimento expressa em função dos elementos da linha de ordem k .

TEOREMA 3.8. DESENVOLVIMENTO POR COMPLEMENTOS ALGÉBRICOS. *Se A'_{kj} é a matriz obtida de A por substituição da linha de ordem k , A_k , pelo vector coordenado unitário I_j , então a fórmula de desenvolvimento escreve-se*

$$(3.19) \quad \det A = \sum_{j=1}^n a_{kj} \det A'_{kj}$$

a qual exprime o determinante de A como uma combinação linear dos elementos da linha de ordem k . O número $\det A'_{kj}$ chama-se o *complemento algébrico* do elemento a_{kj} .

Num teorema que apresentamos a seguir provaremos que cada complemento algébrico é, a menos do sinal, igual ao determinante de uma certa matriz de ordem $n-1$. Estas matrizes designam-se por *menores ou menores complementares* (correspondentes aos elementos da linha de ordem k).

DEFINIÇÃO. *Dada uma matriz quadrada A de ordem $n \geq 2$, a matriz quadrada de ordem $n-1$ obtida da anterior por supressão da linha de ordem k e coluna de ordem j chama-se a menor k, j de A e representa-se por A_{kj} .*

EXEMPLO. Uma matriz $A = (a_{kj})$ de ordem 3 tem nove menores. Três dentre eles são

$$A_{11} = \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}, \quad A_{12} = \begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix}, \quad A_{13} = \begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}.$$

A igualdade (3.2) exprime o determinante de uma matriz 3×3 como uma combinação linear dos determinantes destes três menores. A fórmula pode escrever-se como segue:

$$\det A = a_{11} \det A_{11} - a_{12} \det A_{12} + a_{13} \det A_{13}.$$

O teorema seguinte generaliza esta fórmula ao caso $n \times n$ com $n \geq 2$.

TEOREMA 3.9. DESENVOLVIMENTO SEGUNDO OS MENORES DA LINHA DE ORDEM k . Para qualquer matriz $n \times n$, A , $n \geq 2$, o complemento algébrico de a_{kj} está relacionado com a matriz menor A_{kj} pela fórmula

$$(3.20) \quad \det A'_{kj} = (-1)^{k+j} \det A_{kj}.$$

Consequentemente o desenvolvimento de $\det A$ em função dos elementos da linha de ordem k é dado por

$$(3.21) \quad \det A = \sum_{j=1}^n (-1)^{k+j} a_{kj} \det A_{kj}.$$

Demonstração. Consideremos em primeiro lugar o caso particular em que $k = j = 1$. A matriz A'_{11} tem a forma

$$A'_{11} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix}.$$

Aplicando as operações elementares do tipo (3) às linhas da matriz, podemos converter em zeros todos os elementos situados a seguir ao 1 na primeira coluna, deixando os restantes inalterados. Por exemplo, se multiplicamos a primeira linha de A'_{11} por $-a_{21}$ e adicionamos o resultado à segunda linha, a nova segunda linha vem $(0, a_{22}, a_{23}, \dots, a_{2n})$. Por meio de uma sucessão de tais operações elementares obtemos uma nova matriz que designaremos por A^0_{11} e que é da forma

$$A^0_{11} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & a_{2n} \\ 0 & a_{32} & \cdots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2} & \cdots & a_{nn} \end{bmatrix}.$$

Uma vez que as operações do tipo (3) deixam o determinante inalterado temos

$$(3.22) \quad \det A^0_{11} = \det A'_{11}.$$

Mas A^0_{11} é uma matriz diagonal por blocos pelo que, devido ao teorema 3.7, teremos

$$\det A^0_{11} = \det A_{11},$$

onde A_{11} é a menor $1, 1$ de A ,

$$A_{11} = \begin{bmatrix} a_{22} & \cdots & a_{2n} \\ a_{32} & \cdots & a_{3n} \\ \vdots & & \vdots \\ a_{n2} & \cdots & a_{nn} \end{bmatrix}.$$

Portanto $\det A'_{11} = \det A_{11}$, estando assim demonstrada (3.20) para $k = j = 1$.

Consideremos em seguida o caso particular $k = 1, j$ qualquer, e provemos que

$$(3.23) \quad \det A'_{1j} = (-1)^{j-1} \det A_{1j}.$$

Uma vez demonstrada (3.23), resulta imediatamente a fórmula mais geral (3.20) devido à matriz A'_{kj} se poder transformar numa matriz da forma B'_{ij} , recorrendo a $k-1$ permutações sucessivas de linhas consecutivas. O determinante muda de sinal de cada vez que permutamos duas linhas, pelo que se tem

$$(3.24) \quad \det A'_{kj} = (-1)^{k-1} \det B'_{ij},$$

onde B é uma matriz $n \times n$ cuja primeira linha é I_j e em que a matriz menor $1, j, B_{1j}$, é A_{kj} . Devido a (3.23), temos

$$\det B'_{1j} = (-1)^{j-1} \det B_{1j} = (-1)^{j-1} \det A_{kj},$$

pelo que (3.24) nos dá

$$\det A'_{kj} = (-1)^{k-1} (-1)^{j-1} \det A_{kj} = (-1)^{k+j} \det A_{kj}.$$

Portanto se provarmos (3.23), demonstramos também (3.20).

Passemos então à demonstração de (3.23). A matriz A'_{1j} tem a forma

$$A'_{1j} = \begin{bmatrix} 0 & \cdots & 1 & \cdots & 0 \\ a_{21} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \cdots & a_{nj} & \cdots & a_{nn} \end{bmatrix}.$$

Mediante as operações elementares do tipo (3) sobre as linhas desta matriz, produzimos uma coluna de zeros a seguir ao elemento 1 e transformamo-la em

$$A^0_{1j} = \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ a_{21} & \cdots & a_{2,j-1} & 0 & a_{2,j+1} & \cdots & a_{2n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & \cdots & a_{n,j-1} & 0 & a_{n,j+1} & \cdots & a_{nn} \end{bmatrix}.$$

Como referimos atrás o determinante não varia, isto é $\det A'_{ij} = \det A^0_{ij}$. A matriz menor $1, j, A_{1j}$, tem a forma

$$A_{1j} = \begin{bmatrix} a_{21} & \cdots & a_{2,j-1} & a_{2,j+1} & \cdots & a_{2n} \\ \vdots & & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{n1} & \cdots & a_{n,j-1} & a_{n,j+1} & \cdots & a_{nn} \end{bmatrix}$$

Considere-se agora $\det A^0_{ij}$ como uma função das $n-1$ linhas de A_{1j} , digamos $\det A^0_{ij} = f(A_{1j})$. A função f satisfaz aos três primeiros axiomas para a função determinante de ordem $n-1$. Deste modo, pelo teorema da unicidade, podemos escrever

$$(3.25) \quad f(A_{1j}) = f(J) \det A_{1j},$$

com J representando a matriz identidade de ordem $n-1$. Concluimos assim que para provar (3.23) devemos demonstrar que $f(J) = (-1)^{j-1}$. Por definição $f(J)$ é o determinante da matriz

$$C = \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 1 & & 0 & 0 & 0 & & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 1 & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \quad \leftarrow \text{linha } j$$

\uparrow
 coluna j

Os elementos situados sobre os segmentos oblíquos são todos iguais a 1. Os restantes elementos não escritos são zeros. Permutando a primeira linha de C sucessivamente com as linhas 2, 3, ..., j , chegamos à matriz identidade I , $n \times n$, depois de $j-1$ transposições. O determinante muda de sinal em cada permutação, pelo que $\det C = (-1)^{j-1}$. Consequentemente $f(J) = (-1)^{j-1}$, o que prova (3.23) e portanto (3.20).

3.13. Existência da função determinante

Vamos recorrer à indução em n , a ordem da matriz, para demonstrar a existência da função determinante de qualquer ordem. Para $n=2$ já provámos que existe uma função determinante. Para $n=1$ admitimos por definição que $\det [a_{11}] = a_{11}$.

Supondo que existe uma função determinante de ordem $n - 1$, parece lógico que uma função determinante de ordem n fosse uma das fórmulas de desenvolvimento do teorema 3.9, por exemplo o desenvolvimento segundo os elementos menores da primeira linha. Porém, torna-se mais fácil verificar os axiomas se usarmos uma fórmula diferente, embora do mesmo tipo, expressa em função dos elementos menores da primeira coluna.

TEOREMA 3.10. *Admita-se a existência de determinantes de ordem $n - 1$. Para uma matriz qualquer $n \times n$, $A = (a_{jk})$, seja f a função definida pela fórmula*

$$(3.26) \quad f(A_1, \dots, A_n) = \sum_{j=1}^n (-1)^{j+1} a_{j1} \det A_{j1}.$$

Então f satisfaz aos quatro axiomas para uma função determinante de ordem n e por conseguinte, por indução, os determinantes de ordem n existem para todo n .

Demonstração. Consideremos cada termo da soma (3.26) como uma função das linhas de A e escrevamos

$$f_j(A_1, \dots, A_n) = (-1)^{j+1} a_{j1} \det A_{j1}.$$

Se verificarmos que cada f_j satisfaz aos Axiomas 1 e 2, o mesmo se verificará com f .

Consideremos o efeito da multiplicação da primeira linha de A por um escalar t . O menor A_{11} não será afectado uma vez que nele não intervém a primeira linha de A . O elemento a_{11} vem multiplicado por t , pelo que se terá

$$f_1(tA_1, A_2, \dots, A_n) = ta_{11} \det A_{11} = tf_1(A_1, \dots, A_n).$$

Se $j > 1$ a primeira linha de cada menor A_{j1} resulta multiplicada por t e o coeficiente a_{j1} não é afectado, pelo que mais uma vez temos

$$f_j(tA_1, A_2, \dots, A_n) = tf_j(A_1, A_2, \dots, A_n).$$

Em conclusão cada f_j é homogénea na primeira linha.

Se a linha de ordem k de A é multiplicada por t , com $k > 1$, o menor A_{k1} não é afectado, mas a_{k1} vem multiplicado por t , pelo que f_k é homogénea na linha k . Se $j \neq k$, o coeficiente a_{j1} não é afectado, mas alguma linha de A_{j1} vem multiplicada por t . Consequentemente cada f_j é homogénea na linha k .

Um raciocínio análogo conduzirá à demonstração de que cada f_j é aditiva relativamente a qualquer linha, pelo que f satisfaz aos Axiomas 1 e 2. Provemos agora que f satisfaz ao Axioma 3', uma versão fraca do Axioma 3. Do teorema 3.1 resulta que f satisfaz ao Axioma 3.

Para verificar que f satisfaz ao Axioma 3', admitamos que duas linhas consecutivas

de A são iguais, isto é $A_k = A_{k+1}$. Então, excepto para as matrizes menores A_k e $A_{k+1,1}$, cada menor A_j tem duas linhas iguais pelo que $\det A_j = 0$. Quer dizer que a soma (3.26) consta unicamente dos dois termos correspondentes a $j = k$ e $j = k + 1$,

$$(3.27) \quad f(A_1, \dots, A_n) = (-1)^{k+1} a_{k1} \det A_{k1} + (-1)^{k+2} a_{k+1,1} \det A_{k+1,1}.$$

Mas $A_{k1} = A_{k+1,1}$ e $a_{k1} = a_{k+1,1}$ já que $A_k = A_{k+1}$. Deste modo, os dois termos em (3.27) diferem unicamente pelo sinal, logo $f(A_1, \dots, A_n) = 0$. Assim se provou que f verifica o Axioma 3'.

Finalmente verifiquemos que f satisfaz ao Axioma 4. Quando $A = I$ temos $a_{11} = 1$ e $a_{ji} = 0$ para $j > 1$. Deste modo A_{11} é a matriz identidade de ordem $n - 1$, pelo que cada termo de (3.26) é zero excepto o primeiro, que é igual a 1. Assim $f(I_1, \dots, I_n) = 1$ e f satisfaz ao Axioma 4.

Na demonstração anterior poderia ter-se utilizado uma função f definida em função das matrizes A_{jk} , menores relativos à coluna k em vez das matrizes A_{ji} , menores relativos à primeira coluna. Com efeito, se fizermos

$$(3.28) \quad f(A_1, \dots, A_n) = \sum_{j=1}^n (-1)^{j+k} a_{jk} \det A_{jk},$$

exactamente o mesmo tipo de demonstração permitirá concluir que esta função f satisfaz aos quatro axiomas de uma função determinante. Visto que as funções determinantes são únicas, as fórmulas de desenvolvimento (3.28) e (3.21) são ambas iguais ao $\det A$.

As fórmulas de desenvolvimento (3.28) não só estabelecem a existência de funções determinantes, mas revelam ainda um novo aspecto da teoria dos determinantes — uma conexão entre as propriedades relativas a linhas e as propriedades relativas a colunas. Esta conexão é analisada na secção seguinte.

3.14. O determinante da matriz transposta

Associada a cada matriz A existe outra matriz chamada a *transposta* de A e que se representa por A^t . As linhas de A^t são colunas de A com a mesma ordem. Por exemplo, se

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad \text{então} \quad A^t = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}.$$

Uma definição rigorosa pode ser dada do modo seguinte:

DEFINIÇÃO DE TRANSPOSTA. A transposta de uma matriz $m \times n$, $A = (a_{ij})_{i,j=1}^{m,n}$, é a matriz, $n \times m$, A^t cujo elemento i, j é a_{ji} .

Embora a transposição possa ser efectuada sobre qualquer matriz rectangular vamos confinar-nos às matrizes quadradas, provando de imediato que a transposição de uma matriz quadrada não altera o seu determinante.

TEOREMA 3.11. Para qualquer matriz A , $n \times n$, tem-se

$$\det A = \det A^t.$$

Demonstração. A demonstração faz-se por indução em n . Para $n = 1$ e $n = 2$ o resultado verifica-se imediatamente. Admitamos, pois, que o teorema é verdadeiro para matrizes de ordem $n - 1$. Seja $A = (a_{ij})$ e $B = A^t = (b_{ij})$. Desenvolvendo $\det A$ segundo os elementos da primeira coluna e $\det B$ segundo os elementos da primeira linha temos

$$\det A = \sum_{j=1}^n (-1)^{j+1} a_{j1} \det A_{j1}, \quad \det B = \sum_{j=1}^n (-1)^{j+1} b_{1j} \det B_{1j}.$$

Mas, considerando a definição de transposta de uma matriz, temos $b_{1j} = a_{j1}$ e $B_{1j} = (A_{j1})^t$. Uma vez que estamos a admitir que o teorema é verdadeiro para matrizes de ordem $n - 1$, temos $\det B_{1j} = \det A_{j1}$. Deste modo as somas anteriores são iguais, termo a termo, pelo que $\det A = \det B$.

3.15. A matriz complementos algébricos

O teorema 3.5 mostrou que se A é não singular então $\det A \neq 0$. O teorema que se segue prova o inverso, isto é, se $\det A \neq 0$, então existe A^{-1} . Além disso dá uma fórmula explícita para expressar A^{-1} em função de uma matriz formada a partir dos complementos algébricos dos elementos de A .

No teorema 3.9 provámos que o complemento algébrico i, j de a_{ij} é igual a $(-1)^{i+j} \det A_{ij}$, sendo A_{ij} a matriz menor i, j de A . Representemos este complemento algébrico por $\text{Cal } a_{ij}$. Então, por definição,

$$\text{Cal } a_{ij} = (-1)^{i+j} \det A_{ij}.$$

DEFINIÇÃO DA MATRIZ COMPLEMENTOS ALGÉBRICOS. A matriz cujo elemento i, j é $\text{Cal } a_{ij}$ diz-se a matriz complementos algébricos de A e representa-se por $\text{Cal } A$. Assim tem-se

$$\text{Cal } A = (\text{Cal } a_{ij})_{i,j=1}^n = ((-1)^{i+j} \det A_{ij})_{i,j=1}^n.$$

Na literatura acerca de matrizes, especialmente nos tratados clássicos, a transposta da matriz complementos algébricos, dizem-se *adjunta* de A . Contudo, a nomenclatura actual reserva-se a palavra *adjunta* para outra coisa completamente distinta, que será tratado na secção 5.8.

O teorema seguinte mostra que o produto de A pela sua adjunta (a transposta da matriz complementos algébricos) é, a menos de um factor escalar, a matriz identidade I .

TEOREMA 3.12. *Para qualquer matriz A , $n \times n$, com $n \geq 2$ tem-se*

$$(3.29) \quad A(\text{Cal } A)^t = (\det A)I.$$

Em particular, se $\det A \neq 0$ a inversa de A existe e é definida por

$$A^{-1} = \frac{1}{\det A} (\text{Cal } A)^t.$$

Demonstração. Usando o teorema 3.9 exprimimos $\det A$ em função dos complementos algébricos referentes aos elementos da linha k pela fórmula

$$(3.30) \quad \det A = \sum_{j=1}^n a_{kj} \text{Cal } a_{kj}.$$

Consideremos k fixo e apliquemos esta relação a uma nova matriz B , cuja linha de ordem i é igual à linha de ordem k de A para algum $i \neq k$, e cujas restantes linhas são as mesmas de A . Então $\det B = 0$ porque são iguais as linhas i e k de B . Exprimindo o $\det B$ por intermédio dos complementos algébricos da linha i temos

$$(3.31) \quad \det B = \sum_{j=1}^n b_{ij} \text{Cal } b_{ij} = 0.$$

Mas porque a linha de ordem i de B é igual à linha de ordem k de A temos

$$b_{ij} = a_{kj} \quad \text{e} \quad \text{Cal } b_{ij} = \text{Cal } a_{ij} \quad \text{para todo } j.$$

Por este motivo (3.31) significa que

$$(3.32) \quad \sum_{j=1}^n a_{kj} \text{Cal } a_{ij} = 0 \quad \text{se } k \neq i.$$

As equações (3.30) e (3.32) podem escrever-se conjuntamente

$$(3.33) \quad \sum_{j=1}^n a_{kj} \text{Cal } a_{ij} = \begin{cases} \det A & \text{se } i = k \\ 0 & \text{se } i \neq k. \end{cases}$$

Mas a soma figurando no primeiro membro de (3.33) é o elemento k, i do produto $A(\text{adj } A)^t$. Consequentemente (3.33) implica (3.29).

Como um colorário directo dos teoremas 3.5 e 3.12 temos a seguinte condição necessária e suficiente para que uma matriz quadrada seja não singular.

TEOREMA 3.13. *Uma matriz quadrada A é não singular se e só se $\det A \neq 0$.*

3.16. Regra de Cramer

O teorema 3.12 pode também utilizar-se para resolver um sistema de equações lineares, desde que a matriz dos coeficientes seja não singular. As fórmulas que se obtém traduzem a chamada *regra de Cramer*, em honra do matemático suíço Gabriel Cramer (1704-1752).

TEOREMA 3.14. REGRA DE CRAMER. *Se num sistema de n equações lineares com n incógnitas x_1, \dots, x_n*

$$(3.34) \quad \sum_{j=1}^n a_{ij}x_j = b_i \quad (i = 1, 2, \dots, n)$$

a matriz dos coeficientes $A = (a_{ij})$ é não singular, existe uma única solução para o sistema definida pelas formulas

$$x_j = \frac{1}{\det A} \sum_{k=1}^n b_k \text{Cal } a_{kj}, \quad \text{para } j = 1, 2, \dots, n.$$

Demonstração. O sistema pode escrever-se na forma matricial através da única equação

$$AX = B,$$

com X e B duas matrizes coluna, $X = \begin{bmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix}$, $B = \begin{bmatrix} b_1 \\ \cdot \\ \cdot \\ \cdot \\ b_n \end{bmatrix}$. Visto A ser não singular,

existe uma única solução X definida por

$$(3.35) \quad X = A^{-1}B = \frac{1}{\det A} (\text{Cal } A)^t B.$$

A fórmula (3.34) obtém-se igualando as componentes em (3.35).

Deve observar-se que a fórmula para x_j em (3.34) pode expressar-se como cociente de dois determinantes

$$x_j = \frac{\det C_j}{\det A},$$

onde C_j é a matriz obtida de A pela substituição da coluna j de A pela matriz coluna B .

3.17. Exercícios

1. Determinar a matriz complementos algébricos de cada uma das seguintes matrizes:

$$(a) \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad (b) \begin{bmatrix} 2 & -1 & 3 \\ 0 & 1 & 1 \\ -1 & -2 & 0 \end{bmatrix}, \quad (c) \begin{bmatrix} 3 & 1 & 2 & 4 \\ 2 & 0 & 5 & 1 \\ 1 & -1 & -2 & 6 \\ -2 & 3 & 2 & 3 \end{bmatrix}.$$

2. Determinar a inversa de cada uma das matrizes não singulares do Exercício 1.

3. Determinar todos os valores de λ para os quais $\lambda I - A$ é singular, se A é igual a

$$(a) \begin{bmatrix} 0 & 3 \\ 2 & -1 \end{bmatrix}, \quad (b) \begin{bmatrix} 1 & 0 & 2 \\ 0 & -1 & -2 \\ 2 & -2 & 0 \end{bmatrix}, \quad (c) \begin{bmatrix} 11 & -2 & 8 \\ 19 & -3 & 14 \\ -8 & 2 & -5 \end{bmatrix}.$$

4. Se A é uma matriz $n \times n$ com $n \geq 2$, provar cada uma das propriedades seguintes da correspondente matriz complementos algébricos:

$$(a) \text{Cal } (A') = (\text{adj } A)'. \quad (b) (\text{adj } A)'A = (\det A)I.$$

$$(c) A(\text{adj } A)' = (\text{adj } A)'A \text{ (} A \text{ permuta com a adjunta).}$$

5. Resolver cada um dos sistemas seguintes, utilizando a regra de Cramer:

$$(a) x + 2y + 3z = 8, \quad 2x - y + 4z = 7, \quad -y + z = 1.$$

$$(b) x + y + 2z = 0, \quad 3x - y - z = 3, \quad 2x + 5y + 3z = 4.$$

6. (a) Justificar que cada uma das equações que se seguem é a equação cartesiana de uma recta, no plano XOY , a qual passa pelos pontos (x_1, y_1) e (x_2, y_2) .

$$\det \begin{bmatrix} x & x_1 & y & y_1 \\ x_2 & x_1 & y_2 & y_1 \end{bmatrix} = 0; \quad \det \begin{bmatrix} x & y & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{bmatrix} = 0.$$

(b) Estabelecer e demonstrar relações análogas para um plano, no espaço tridimensional, passando por três pontos não colineares.

(c) Estabelecer e demonstrar relações análogas para uma circunferência, no plano XOY , que passa por três pontos não colineares.

7. Dadas n^2 funções f_{ij} cada uma das quais derivável no intervalo (a, b) , definir $F(x) = \det [f_{ij}(x)]$ para cada x em (a, b) . Provar que a derivada $F'(x)$ é uma soma de n determinantes

$$F'(x) = \sum_{i=1}^n \det A_i(x),$$

onde $A_i(x)$ é a matriz obtida por derivação das funções da linha i de $[f_{ij}(x)]$.

8. Uma matriz $n \times n$ de funções da forma $W(x) = [u^{(j-1)}(x)]$, na qual cada linha depois da primeira é a derivada da linha anterior, chama-se a *matriz Wronskiana*, em homenagem ao matemático polaco J. M. H. Wronski (1778-1853). Provar que a derivada do determinante de $W(x)$ é o determinante da matriz obtida derivando cada um dos elementos da última linha de $W(x)$. [Sugestão: utilizar o Exercício 7.]

4

VALORES PRÓPRIOS E VECTORES PRÓPRIOS

4.1. Transformações lineares representadas por matrizes diagonais

Seja $T: V \rightarrow V$ uma transformação linear definida num espaço linear V de dimensão finita. As propriedades de T , que são independentes de qualquer sistema de coordenadas (base) definido em V , dizem-se *propriedades intrínsecas* de T . Elas estarão presentes em todas as representações matriciais de T . Se puder ser escolhida uma base de maneira que a matriz daí resultante tenha uma forma particularmente simples, poderá tornar-se possível a detecção de certas propriedades intrínsecas directamente a partir de representação matricial.

Entre os mais simples tipos de matrizes estão as matrizes diagonais. Portanto será de pôr a questão de se saber se cada transformação linear pode representar-se mediante uma matriz diagonal. No capítulo 2 tratámos do problema de determinação de uma representação da transformação linear $T: V \rightarrow W$ por uma matriz diagonal, com $\dim V = n$ e $\dim W = m$. No teorema 2.14 provámos que existe sempre uma base (e_1, \dots, e_n) para V e uma base (w_1, \dots, w_m) para W , tais que a matriz de T relativa a este par de bases é uma matriz diagonal. Em particular, se $W = V$ a matriz será uma matriz diagonal quadrada. O aspecto novo do problema reside agora no facto de pretendermos usar a mesma base para V e W . Com esta restrição nem sempre é possível determinar uma matriz diagonal para a representação de T . Voltamos, então, ao problema da determinação daquelas transformações que admitem uma representação por uma matriz diagonal.

Notação: Se $A = (a_{ij})$ é uma matriz diagonal, escrevemos $A = \text{diag } (a_{11}, a_{22}, \dots, a_{nn})$.

É fácil estabelecer uma condição necessária e suficiente para que uma transformação linear tenha uma representação matricial diagonal.

TEOREMA 4.1. *Seja $T: V \rightarrow V$ uma transformação linear, com $\dim V = n$. Se T admite uma representação matricial diagonal, então existe em V um conjunto independente de elementos u_1, \dots, u_n e um correspondente conjunto de escalares $\lambda_1, \dots, \lambda_n$ tais que*

$$(4.1) \quad T(u_k) = \lambda_k u_k \quad \text{para } k = 1, 2, \dots, n.$$

Inversamente, se existe um conjunto independente u_1, \dots, u_n de elementos de V e um correspondente conjunto de escalares $\lambda_1, \dots, \lambda_n$ verificando (4.1), então a matriz

$$A = \text{diag}(\lambda_1, \dots, \lambda_n)$$

é uma representação de T , relativa à base (u_1, \dots, u_n) .

Demonstração. Admitamos em primeiro lugar que T admite uma representação matricial diagonal, $A = (a_{ij})$, relativa a determinada base (e_1, \dots, e_n) . A acção de T sobre os elementos da base é dada pela fórmula

$$T(e_k) = \sum_{i=1}^n a_{ik} e_i = a_{kk} e_k$$

já que $a_{ik} = 0$ para $i \neq k$. Isto demonstra (4.1) com $u_k = e_k$ e $\lambda_k = a_{kk}$.

Admitamos agora que existem os elementos independentes u_1, \dots, u_n e os escalares $\lambda_1, \dots, \lambda_n$ verificando (4.1). Visto que u_1, \dots, u_n são independentes, constituem uma base para V . Se definirmos $a_{kk} = \lambda_k$ e $a_{ik} = 0$ para $i \neq k$, então a matriz $A = (a_{ik})$ é uma matriz diagonal que representa T em relação à base (u_1, \dots, u_n) .

Assim o problema de obtenção de uma representação matricial diagonal para uma transformação linear foi transformado num outro problema, o da determinação de elementos independentes u_1, \dots, u_n e escalares $\lambda_1, \dots, \lambda_n$ verificando (4.1). Os elementos u_k e os escalares λ_k verificando (4.1) dizem-se *vectores próprios* e *valores próprios* de T , respectivamente. Estudaremos, na secção seguinte, os valores próprios e vectores próprios duma maneira mais geral.

4.2. Valores próprios e vectores próprios de uma transformação linear

Nesta discussão V representa um espaço linear e S um subespaço de V . Os espaços S e V não são necessariamente de dimensão finita.

DEFINIÇÃO. *Seja $T: S \rightarrow V$ uma transformação linear de S em V . Um escalar λ diz-se um valor próprio de T se existir em S um elemento x não nulo tal que*

$$(4.2) \quad T(x) = \lambda x.$$

o elemento x chama-se um vector próprio de T correspondente a λ . O escalar λ diz-se um valor próprio respeitante a x .

Existe precisamente um valor próprio dizendo respeito a um vector próprio dado x . Com efeito, se tivermos $T(x) = \lambda x$ e $T(x) = \mu x$ para um certo $x \neq 0$ para um certo $x \neq 0$, então $\lambda x = \mu x$ donde resulta $\lambda = \mu$.

Nota: Embora a equação (4.2) seja sempre válida para $x = 0$ e qualquer escalar λ , a definição exclui 0 como vector próprio. Uma razão do facto da exclusão do 0 como vector próprio, reside em que deverá existir um único valor próprio λ associado a um vector próprio dado x .

Os exemplos que se seguem classificam o significado destes conceitos.

EXEMPLO 1. *Multiplicação por um escalar dado.* Seja $T: S \rightarrow V$ a transformação linear definida pela equação $T(x) = cx$ para todo x de S , sendo c um escalar determinado. Neste exemplo todo o elemento não nulo de S é um vector próprio correspondente ao escalar c .

EXEMPLO 2. *O espaço $E(\lambda)$ formado por todos os elementos x tais que $T(x) = \lambda x$.* Seja $T: S \rightarrow V$ uma transformação linear que admite o valor próprio λ . Seja $E(\lambda)$ o conjunto de todos os elementos x de S tais que $T(x) = \lambda x$. Este conjunto contém o elemento zero 0 e todos os vectores próprios correspondentes a λ . É fácil provar que $E(\lambda)$ é um subespaço de S , porque se x e y pertencem a $E(\lambda)$ se tem

$$T(ax + by) = aT(x) + bT(y) = a\lambda x + b\lambda y = \lambda(ax + by)$$

para escalares a e b quaisquer. Por conseguinte $(ax + by) \in E(\lambda)$, pelo que $E(\lambda)$ é um subespaço. O espaço $E(\lambda)$ chama-se o *espaço invariante* correspondente a λ . Pode ser de dimensão finita ou infinita. Se $E(\lambda)$ tem dimensão finita, então $\dim E(\lambda) \geq 1$, visto que $E(\lambda)$ contém pelo menos um elemento não nulo, x , correspondente a λ .

EXEMPLO 3. *Existência de valores próprios nulos.* Se existe um vector próprio, ele não pode ser nulo por definição. Contudo, o escalar zero *pode ser* um valor próprio. Com efeito, se 0 é um valor próprio para x então $T(x) = 0x = 0$, pelo que x pertence ao espaço nulo de T . Inversamente, se o espaço nulo de T contém qualquer elemento não nulo então cada um deles é um vector próprio correspondente ao valor próprio 0 . Em geral, $E(0)$ é o espaço nulo de $T - 0I$.

EXEMPLO 4. *Simetria no plano XOY.* Seja $S = V = V_1(\mathbb{R})$ e seja T uma reflexão no plano XOY, isto é, T actua sobre i, j, k do modo seguinte: $T(i) = i$, $T(j) = j$, $T(k) = -k$. Cada vector não nulo no plano XOY é um vector próprio com valor próprio 1 . Os restantes vectores próprios são os da forma ck , com $c \neq 0$; cada um deles tem valor próprio -1 .

EXEMPLO 5. *Rotação de um ângulo α no plano e em torno de um ponto fixo.* Este exemplo é de interesse especial porque mostra que a existência de vectores próprios pode

depende do corpo fundamental de escalares. O plano pode considerar-se como um espaço linear de duas maneiras diferentes: (1) como um espaço linear *real* bidimensional, $V = V_2(\mathbf{R})$ com dois elementos base $(1, 0)$, $(0, 1)$ e com os números reais como escalares; (2) como um espaço linear *complexo* unidimensional, $V = V_1(\mathbf{C})$, com um elemento base 1 e números complexos como escalares.

Consideremos em primeiro lugar a segunda interpretação. Cada elemento $z \neq 0$ de $V_1(\mathbf{C})$ pode exprimir-se na forma polar, $z = re^{i\theta}$. Se T imprime a z uma rotação de um ângulo α , então $T(z) = re^{i(\theta + \alpha)} = e^{i\alpha}z$. Assim, cada $z \neq 0$ é um vector próprio com valor próprio $\lambda = e^{i\alpha}$. Observe-se que o valor $e^{i\alpha}$ não é real a menos que α seja um múltiplo inteiro de π .

Consideremos agora o plano como um espaço linear *real*, $V_2(\mathbf{R})$. Porque os escalares de $V_2(\mathbf{R})$ são números reais, a rotação T tem valores próprios reais somente se α é um inteiro múltiplo de π . Por outras palavras, se α não for um inteiro múltiplo de π , então T não admite valores próprios reais e, por conseguinte, não tem valores próprios. Quer dizer que a existência de vectores próprios e valores próprios pode depender da escolha dos escalares para V .

EXEMPLO 6. *O operador derivação.* Seja V o espaço linear de todas as funções reais f admitindo derivadas de todas as ordens num dado intervalo aberto. Seja D a transformação linear que aplica cada f sobre a sua derivada, $D(f) = f'$. Os vectores próprios de D são as funções não nulas f satisfazendo a uma equação da forma

$$f' = \lambda f$$

para algum real λ . Esta é uma equação diferencial linear de primeira ordem. Todas as suas soluções são dadas pela fórmula

$$f(x) = ce^{\lambda x},$$

onde c é uma constante arbitrária real. Deste modo os vectores próprios de D são todas as funções exponenciais $f(x) = ce^{\lambda x}$ com $c \neq 0$. O valor próprio correspondente a $f(x) = ce^{\lambda x}$ é λ . Em exemplos como este em que V é um espaço funcional os vectores próprios chamam-se *funções próprias*.

EXEMPLO 7. *O operador integração.* Seja V o espaço linear de todas as funções reais contínuas num intervalo finito $[a, b]$. Se $f \in V$, definamos $g = T(f)$ como sendo a função dada por

$$g(x) = \int_a^x f(t) dt \quad \text{se } a \leq x \leq b.$$

As funções próprias de T (se existirem) são as funções não nulas f satisfazendo a uma equação da forma

$$(4.3) \quad \int_a^x f(t) dt = \lambda f(x)$$

para algum real λ . Se existe uma função própria podemos derivar esta igualdade para obter a relação $f(x) = \lambda f'(x)$, a partir da qual obtemos $f(x) = ce^{x/\lambda}$, desde que $\lambda \neq 0$. Por outras palavras, as únicas funções próprias possíveis são as funções exponenciais da forma $f(x) = ce^{x/\lambda}$, com $c \neq 0$ e $\lambda \neq 0$. Contudo, se fazemos $x = a$ em (4.3) resulta

$$0 = \lambda f(a) = \lambda ce^{a/\lambda}.$$

Mas porque $e^{a/\lambda}$ nunca é zero, vemos que a equação $T(f) = \lambda f$ não pode ser satisfeita com uma função f não nula, pelo que T não tem nem funções próprias nem valores próprios.

EXEMPLO 8. *O subespaço gerado por um vector próprio.* Seja $T: S \rightarrow V$ uma transformação linear admitindo um valor próprio λ . Seja x um vector próprio correspondente a λ e seja $L(x)$ o subespaço gerado por x , isto é, $L(x)$ é o conjunto de todos os produtos de escalares por x . É fácil mostrar que T aplica $L(x)$ em si próprio. Com efeito, se $y = cx$ temos

$$T(y) = T(cx) = cT(x) = c(\lambda x) = \lambda(cx) = \lambda y.$$

Se $c \neq 0$ então $y \neq 0$ pelo que cada elemento y de $L(x)$ não nulo é também um vector próprio correspondente a λ .

Um subespaço U de S diz-se *invariante* sob a transformação T se T aplica cada elemento de U sobre um elemento de U . Acabámos de provar que o subespaço gerado por um vector próprio é invariante sob T .

4.3. Independência linear de vectores próprios correspondentes a valores próprios distintos

Uma das propriedades mais importantes dos valores próprios é referida no teorema que se segue. Como atrás, S representa um subespaço de um espaço linear V .

TEOREMA 4.2. *Se u_1, \dots, u_k são vectores próprios de uma transformação linear $T: S \rightarrow V$ e os correspondentes valores próprios $\lambda_1, \dots, \lambda_k$ são distintos, então os vectores próprios u_1, \dots, u_k são independentes.*

Demonstração. A demonstração faz-se por indução em k . O resultado é trivial quando $k = 1$. Suponhamos, então, que o teorema se demonstrou para qualquer conjunto de $k-1$ vectores próprios. Sejam u_1, \dots, u_k os k vectores próprios correspondentes a valores próprios distintos, e admitamos que existem escalares c_i tais que

$$(4.4) \quad \sum_{i=1}^k c_i u_i = 0.$$

Aplicando T a ambos os membros de (4.4) e usando o facto de que $T(u_i) = \lambda_i u_i$ encontramos

$$(4.5) \quad \sum_{i=1}^k c_i \lambda_i u_i = 0.$$

Multiplicando (4.4) por λ_k e subtraindo de (4.5), obtemos a equação

$$\sum_{i=1}^{k-1} c_i (\lambda_i - \lambda_k) u_i = 0.$$

Mas porque u_1, \dots, u_{k-1} são independentes, deve verifica-se $c_i (\lambda_i - \lambda_k) = 0$ para cada $i = 1, 2, \dots, k-1$. Uma vez que os valores próprios são distintos temos $\lambda_i \neq \lambda_k$ para $i \neq k$, o que implica $c_i = 0$ para $i = 1, 2, \dots, k-1$. De (4.4) vemos que c_k é também 0 pelo que os vectores próprios u_1, \dots, u_k são independentes.

Observe-se que o teorema 4.2 não seria verdadeiro se o elemento zero pudesse ser um vector próprio. Esta é outra razão para a exclusão do O como vector próprio.

Observação. O inverso do teorema 4.2 não é verdadeiro, isto é, se T admite os vectores próprios independentes u_1, \dots, u_k , então os valores próprios que en dizem respeito, $\lambda_1, \dots, \lambda_k$, não são necessariamente distintos. Por exemplo, se T é a transformação identidade, $T(x) = x$ para qualquer x , então cada $x \neq O$ é um vector próprio mas existe unicamente um valor próprio, $\lambda = 1$.

O teorema 4.2. tem consequências importantes para o caso de espaços de dimensão finita.

TEOREMA 4.3. *Se $\dim V = n$, toda transformação linear $T: V \rightarrow V$ admite quando muito n valores próprios distintos. Se T tem precisamente n valores próprios distintos, então os correspondentes vectores próprios formam uma base para V e a matriz de T , me relação a esta base, é uma matriz diagonal com os valores próprios como elementos diagonais.*

Demonstração. Se existissem $n + 1$ valores próprios distintos então, pelo teorema 4.2, V conteria $n + 1$ elementos independentes. Tal não é possível porque $\dim V = n$. A segunda afirmação resulta dos teoremas 4.1 e 4.2.

Nota: O teorema 4.3 diz-nos a existência de n valores próprios distintos é uma condição *suficiente* para T admitir uma representação matricial diagonal. Esta condição não é, porém, necessária. Existem transformações lineares com menos do que n valores próprios, as quais podem ser representadas por matrizes diagonais. A transformação identidade é um exemplo. Todos os seus valores próprios são iguais a 1, mas pode representar-se pela matriz identidade. O teorema 4.1 diz-nos que a existência de n vectores próprios independentes é uma condição *necessária* e *suficiente* para que T tenha uma representação matricial diagonal.

4.4. Exercícios

- (a) Se T admite um valor próprio λ , provar que aT admite o valor próprio $a\lambda$.
 (b) Se x é um vector próprio para T_1 e T_2 , provar que x é também um vector próprio para $aT_1 + bT_2$. Como estão relacionados os valores próprios?
- Suponhamos que $T: V \rightarrow V$ tem um vector próprio x correspondente ao valor próprio λ . Provar que x é um vector próprio de T^2 correspondendo a λ^2 e, mais geralmente, x é um vector próprio de T^n correspondente de λ^n . Utilizar depois o resultado do Exercício 1 para mostrar que se P é um polinómio, então x é um vector próprio de $P(T)$ correspondente a $P(\lambda)$.
- Consideremos o plano como um espaço linear real, $V = V_2(\mathbb{R})$, e seja T uma rotação de V de $\pi/2$ radianos. Embora T não possua vectores próprios, provar que todo o vector não nulo é um vector próprio de T^2 .
- Se $T: V \rightarrow V$ goza da propriedade de T^2 admitir um valor próprio não negativo λ^2 , provar que pelo menos um dos valores λ ou $-\lambda$ é um valor próprio para T . [Sugestão: $T^2 - \lambda^2 I = (T + \lambda I)(T - \lambda I)$.]
- Seja V o espaço linear de todas as funções reais deriváveis no intervalo $(0, 1)$. Se $f \in V$, definir $g = T(f)$, pondo $g(t) = tf'(t)$ para todo t em $(0, 1)$. Provar que cada real λ é um valor próprio para T , e determinar as funções próprias correspondentes a λ .
- Seja V o espaço linear de todos os polinómios reais $p(x)$ de grau $\leq n$. Se $p \in V$, definir $q = T(p)$ significando $q(t) = p(t+1)$ para todo o real t . Provar que T admite unicamente o valor próprio 1. Quais são as funções próprias correspondentes a este valor próprio?
- Seja V o espaço linear de todas as funções contínuas em $(-\infty, +\infty)$ e tais que o integral $\int_x^\infty f(t)dt$ existe para todo o real x . Se $f \in V$, seja $g = T(f)$ definida pela equação $g(x) = \int_{-\infty}^x f(t)dt$. Provar que cada positivo λ é um valor próprio para T e determinar as funções próprias correspondentes a λ .
- Seja V o espaço linear de todas as funções contínuas em $(-\infty, +\infty)$ e tais que o integral $\int_{-\infty}^x tf(t)dt$ exista para todo o real x . Se $f \in V$, seja $g = T(f)$ definida pela equação $g(x) = \int_{-\infty}^x tf(t)dt$. Provar que cada λ negativo é um valor próprio para T e determinar as funções próprias correspondentes a λ .
- Seja $V = C(0, \pi)$ o espaço linear real de todas as funções reais contínuas no intervalo $[0, \pi]$. Seja S o subespaço de todas as funções f as quais têm derivadas de segunda ordem contínuas e que verificam as condições fronteira $f(0) = f(\pi) = 0$. Seja $T: S \rightarrow V$ a transformação linear que aplica cada f sobre a sua segunda derivada, $T(f) = f''$. Provar que os valores próprios de T são números da forma $-n^2$, onde $n = 1, 2, \dots$, e que as funções próprias correspondentes a $-n^2$ são $f(t) = c_n \sin nt$, onde $c_n \neq 0$.
- Seja V o espaço linear de todas as sucessões reais convergentes $\{x_n\}$. Definase $T: V \rightarrow V$ do modo seguinte: Se $x = \{x_n\}$ é uma sucessão convergente com limite a , seja $T(x) = \{y_n\}$, onde $y_n = a - x_n$ para $n \geq 1$. Provar que T admite unicamente dois valores próprios, $\lambda = 0$ e $\lambda = -1$, e determinar os vectores próprios correspondentes a cada λ .
- Admitir que uma transformação linear T tem dois vectores próprios x e y correspondendo a valores próprios distintos λ e μ . Se $ax + by$ é um vector próprio de T , provar que $a = 0$ ou $b = 0$.
- Seja $T: S \rightarrow V$ uma transformação linear tal que cada elemento não nulo de S seja um vector próprio. Provar que existe um escalar c tal que $T(x) = cx$. Por outras palavras, a única transformação com esta propriedade é o produto da identidade por um escalar. [Sugestão: Fazer uso do Exercício 11.]

4.5. O caso de dimensão finita. Polinómios característicos

Se $\dim V = n$, o problema de determinação de valores próprios de uma transformação linear $T: V \rightarrow V$ pode ser resolvido com recurso aos determinantes. Pretendemos determinar aqueles escalares λ tais que a equação $T(x) = \lambda x$ tenha uma solução com $x \neq 0$. A equação $T(x) = \lambda x$ pode escrever-se na forma

$$(\lambda I - T)(x) = 0,$$

sendo I a transformação identidade. Se fizermos $T_\lambda = \lambda I - T$, então λ é um valor próprio se e somente se a equação

$$(4.6) \quad T_\lambda(x) = 0$$

tem uma solução x não nula, caso em que T_λ é não invertível (devido ao teorema 2.10). Portanto, pelo teorema 2.20, uma solução não nula de (4.6) existe se e só se a matriz de T_λ é singular. Se A é uma representação matricial de T , então $\lambda I - A$ é uma representação matricial para T_λ . Pelo teorema 3.13, a matriz $\lambda I - A$ é singular se e só se $\det(\lambda I - A) = 0$. Quer dizer, se λ é um valor próprio para T , verifica a equação

$$(4.7) \quad \det(\lambda I - A) = 0.$$

Inversamente, qualquer λ do corpo fundamental de escalares que satisfaz a (4.7) é um valor próprio. Isto sugere o estudo do determinante $\det(\lambda I - A)$ considerado como uma função de λ .

TEOREMA 4.4. *Se A é uma matriz $n \times n$ e se I é a matriz identidade $n \times n$, a função f definida pela equação*

$$f(\lambda) = \det(\lambda I - A)$$

é um polinómio em λ de grau n . Além disso, o termo de maior grau é λ^n e o termo de grau zero é $f(0) = \det(-A) = (-1)^n \det A$.

Demonstração. A afirmação $f(0) = \det(-A)$ resulta imediatamente da definição de f . Vamos provar que f é um polinómio de grau n unicamente para o caso $n \leq 3$. A demonstração para o caso geral pode ser feita por indução e é deixada como exercício. (Ver Exercício 9 da Secção 4.8.)

Para $n = 1$ o determinante é o polinómio linear $f(\lambda) = \lambda - a_{11}$. Para $n = 2$ temos

$$\begin{aligned} \det(\lambda I - A) &= \begin{vmatrix} \lambda - a_{11} & -a_{12} \\ -a_{21} & \lambda - a_{22} \end{vmatrix} = (\lambda - a_{11})(\lambda - a_{22}) - a_{12}a_{21} \\ &= \lambda^2 - (a_{11} + a_{22})\lambda + (a_{11}a_{22} - a_{12}a_{21}), \end{aligned}$$

um polinómio quadrático em λ . Para $n = 3$ temos

$$\det(\lambda I - A) = \begin{vmatrix} \lambda - a_{11} & -a_{12} & -a_{13} \\ -a_{21} & \lambda - a_{22} & -a_{23} \\ -a_{31} & -a_{32} & \lambda - a_{33} \end{vmatrix}$$

$$= (\lambda - a_{11}) \begin{vmatrix} \lambda - a_{22} & -a_{23} \\ -a_{32} & \lambda - a_{33} \end{vmatrix} + a_{12} \begin{vmatrix} -a_{21} & -a_{23} \\ -a_{31} & \lambda - a_{33} \end{vmatrix} - a_{13} \begin{vmatrix} -a_{21} & \lambda - a_{22} \\ -a_{31} & -a_{32} \end{vmatrix}.$$

Os dois últimos termos são polinómios lineares em λ . O primeiro termo é um polinómio do 3.º grau, sendo o termo de maior grau λ^3 .

DEFINIÇÃO. Se A é uma matriz $n \times n$, o determinante

$$f(\lambda) = \det(\lambda I - A)$$

chama-se o polinómio característico de A .

As raízes do polinómio característico de A são números complexos, podendo algumas delas serem reais. Se F representa quer o campo real \mathbb{R} quer o campo complexo \mathbb{C} , temos o seguinte teorema:

TEOREMA 4.5. Seja $T: V \rightarrow V$ uma transformação linear, com F o corpo dos escalares de V e $\dim V = n$. Seja A a representação matricial de T . O conjunto dos valores próprios de T é formado pelas raízes do polinómio de A que pertencem a F .

Demonstração. A discussão do teorema 4.4 mostra que cada valor próprio de T satisfaz à equação $\det(\lambda I - A) = 0$ e que qualquer raiz do polinómio característico de A que pertence a F é um valor próprio de T .

A matriz A depende de escolha da base em V , mas os valores próprios de T foram definidos sem referência a uma base dada. Portanto, o conjunto das raízes do polinómio característico de A deve ser independente da escolha da base. Mas pode ir-se mais longe, pois na Secção seguinte demonstraremos que o próprio polinómio característico é independente da escolha da base. Retomamos agora o problema do cálculo dos valores próprios e vectores próprios no caso de espaços de dimensão finita.

4.6. Cálculo de valores próprios e vectores próprios no caso de dimensão finita

No caso de dimensão finita os valores próprios e vectores próprios de uma transformação linear T chamam-se também valores próprios e vectores próprios de cada representação matricial de T . Assim, os valores próprios da matriz quadrada A são as raízes do polinómio característico $f(\lambda) = \det(\lambda I - A)$. Os vectores próprios correspondentes a um valor próprio λ são os vectores não nulos $X = (x_1, \dots, x_n)$ considerados como matrizes coluna $n \times 1$ que satisfazem a equação matricial.

$$AX = \lambda X, \quad \text{ou} \quad (\lambda I - A)X = 0.$$

Este é um sistema de n equações lineares para as componentes x_1, \dots, x_n . Uma vez conhecido λ podemos obter os vectores próprios resolvendo esse sistema. Esclarecemos o assunto com três exemplos que apresentam aspectos diferentes.

EXEMPLO 1. *Uma matriz com todos os valores próprios distintos. A matriz*

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 2 & 3 & 4 \\ -1 & -1 & -2 \end{bmatrix}$$

tem o polinómio característico

$$\det(\lambda I - A) = \det \begin{bmatrix} \lambda - 2 & -1 & -1 \\ -2 & \lambda - 3 & -4 \\ 1 & 1 & \lambda + 2 \end{bmatrix} = (\lambda - 1)(\lambda + 1)(\lambda - 3),$$

pelo que os três valores próprios são distintos: 1, -1, e 3. Para determinar os vectores próprios correspondentes a $\lambda = 1$, resolvemos o sistema $AX = X$, ou

$$\begin{bmatrix} 2 & 1 & 1 \\ 2 & 3 & 4 \\ -1 & -1 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

que nos dá

$$2x_1 + x_2 + x_3 = x_1$$

$$2x_1 + 3x_2 + 4x_3 = x_2$$

$$-x_1 - x_2 - 2x_3 = x_3,$$

o qual pode escrever-se

$$x_1 + x_2 + x_3 = 0$$

$$2x_1 + 2x_2 + 4x_3 = 0$$

$$-x_1 - x_2 - 3x_3 = 0.$$

Adicionando membro a membro a primeira e terceira equações encontramos $x_3 = 0$ e as três equações reduzem-se a $x_1 + x_2 = 0$. Deste modo os vectores próprios correspondentes a $\lambda = 1$ são $X = t(1, -1, 0)$, com t qualquer escalar não nulo.

Por cálculos análogos encontramos os vectores próprios $X = t(0, 1, -1)$ correspondentes a $\lambda = -1$, e $X = t(2, 3, -1)$ correspondentes a $\lambda = 3$, com t qualquer escalar

não nulo. Uma vez que os valores próprios são distintos os correspondentes vectores próprios $(1, -1, 0)$, $(0, 1, -1)$ e $(2, 3, -1)$ são independentes. Os resultados podem resumir-se da maneira seguinte, indicando-se na terceira coluna a dimensão do espaço invariante $E(\lambda)$.

Valor próprio λ	Vectores próprios	$\dim E(\lambda)$
1	$t(1, -1, 0), \quad t \neq 0$	1
-1	$t(0, 1, -1), \quad t \neq 0$	1
3	$t(2, 3, -1), \quad t \neq 0$	1

EXEMPLO 2. Uma matriz com valores próprios repetidos. A matriz

$$A = \begin{bmatrix} 2 & -1 & 1 \\ 0 & 3 & -1 \\ 2 & 1 & 3 \end{bmatrix}$$

corresponde o polinómio característico

$$\det(\lambda I - A) = \det \begin{bmatrix} \lambda - 2 & 1 & -1 \\ 0 & \lambda - 3 & 1 \\ -2 & -1 & \lambda - 3 \end{bmatrix} = (\lambda - 2)(\lambda - 2)(\lambda - 4).$$

Os valores próprios são 2, 2 e 4. (Citamos o valor próprio 2 duas vezes para evidenciar que é uma raiz dupla do polinómio característico.) Para determinar os vectores próprios correspondentes a $\lambda = 2$ resolvemos o sistema $AX = 2X$, que se reduz a

$$-x_2 + x_3 = 0$$

$$x_2 - x_3 = 0$$

$$2x_1 + x_2 + x_3 = 0.$$

Este tem a solução $x_2 = x_3 = -x_1$ pelo que os vectores próprios correspondentes a $\lambda = 2$ são $t(-1, 1, 1)$, com $t \neq 0$. Analogamente encontramos $t(1, -1, 1)$ correspondendo ao valor próprio $\lambda = 4$. Os resultados podem resumir-se como segue:

Valor próprio λ	Vectores próprios	$\dim E(\lambda)$
2, 2	$t(-1, 1, 1), \quad t \neq 0$	1
4	$t(1, -1, 1), \quad t \neq 0$	1

EXEMPLO 3. Outra matriz com valores próprios repetidos. A matriz

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 2 & 3 & 2 \\ 3 & 3 & 4 \end{bmatrix}$$

admite o polinómio característico $(\lambda - 1)(\lambda - 1)(\lambda - 7)$. Quando $\lambda = 7$ o sistema $AX = 7X$ vem

$$\begin{aligned} 5x_1 - x_2 - x_3 &= 0 \\ -2x_1 + 4x_2 - 2x_3 &= 0 \\ -3x_1 - 3x_2 + 3x_3 &= 0. \end{aligned}$$

Este admite a solução $x_2 = 2x_1$, $x_3 = 3x_1$, pelo que os vectores próprios correspondentes a $\lambda = 7$ são $t(1, 2, 3)$, com $t \neq 0$. Para o valor próprio $\lambda = 1$, o sistema $AX = X$ consta da equação

$$x_1 + x_2 + x_3 = 0$$

repetida três vezes. Para resolver esta equação podemos fazer $x_1 = a$, $x_2 = b$, com a e b arbitrários e tomar $x_3 = -a - b$. Assim cada vector próprio correspondente a $\lambda = 1$ tem a forma

$$(a, b, -a - b) = a(1, 0, -1) + b(0, 1, -1),$$

onde $a \neq 0$ e $b \neq 0$. Isto significa que os vectores $(1, 0, -1)$ e $(0, 1, -1)$ forma uma base para $E(1)$. Consequentemente $E(\lambda) = 2$ quando $\lambda = 1$. Os resultados podem resumir-se como segue.

<u>Valor próprio</u>	<u>Vectores próprios</u>	<u>dim $E(\lambda)$</u>
7	$t(1, 2, 3), \quad t \neq 0$	1
1, 1	$a(1, 0, -1) + b(0, 1, -1), \quad a, b \text{ não ambos nulos}$	2

Observe-se que neste exemplo existem três vectores próprios independentes, mas somente dois valores próprios distintos.

4.7. Traço de uma matriz

Seja $f(\lambda)$ o polinómio característico de uma matriz A , $n \times n$. Representemos as n raízes de $f(\lambda)$ por $\lambda_1, \dots, \lambda_n$, com cada raiz escrita tantas vezes quantas as unidades que o seu grau de multiplicidade indica. Então pelo teorema de factorização tem-se

$$f(\lambda) = (\lambda - \lambda_1) \cdots (\lambda - \lambda_n).$$

Podemos também escrever $f(\lambda)$ ordenado segundo as potências decrescentes de λ na forma

$$f(\lambda) = \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0.$$

Comparando esta com a forma factorizada concluímos que o termo constante c_0 e o coeficiente de λ^{n-1} são dados pelas fórmulas

$$c_0 = (-1)^n \lambda_1 \cdots \lambda_n \quad \text{e} \quad c_{n-1} = -(\lambda_1 + \cdots + \lambda_n).$$

Visto ser também $c_0 = (-1)^n \det A$, vê-se que

$$\lambda_1 \cdots \lambda_n = \det A.$$

isto é, o produto das raízes do polinómio característico de A é igual ao determinante de A .

A soma das raízes de $f(\lambda)$ chama-se o traço de A , e representa-se por $\text{tr } A$. Desta maneira temos, por definição,

$$\text{tr } A = \sum_{i=1}^n \lambda_i.$$

O coeficiente de λ^{n-1} é dado por $c_{n-1} = -\text{tr } A$. Podemos também calcular este coeficiente a partir de $f(\lambda)$ na forma de determinante e encontramos

$$c_{n-1} = -(a_{11} + \cdots + a_{nn}).$$

(No Exercício 12 da Secção 4.8 pede-se uma demonstração desta fórmula.) As duas fórmulas para c_{n-1} mostram que

$$\text{tr } A = \sum_{i=1}^n a_{ii}.$$

Isto é, o traço de A é também igual à soma dos elementos diagonais de A .

Uma vez que a soma dos elementos diagonais é fácil de calcular, ela pode usar-se como verificação numérica no cálculo de valores próprios. Nas secções que se seguem analisaremos outras propriedades do traço de uma matriz.

4.8. Exercícios

Determinar os valores próprios e vectores próprios de cada uma das matrizes dos Exercícios 1 a 3. Determinar igualmente, para cada λ , a dimensão do espaço $E(\lambda)$.

1. (a) $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, (b) $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, (c) $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$, (d) $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$.

2. $\begin{bmatrix} 1 & a \\ b & 1 \end{bmatrix}$, $a > 0$, $b > 0$.

3. $\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$.

4. As matrizes $P_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $P_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$, $P_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$

aparecem na teoria quântica e dizem-se *matrizes spin de Pauli*, em honra do físico Wolfgang Pauli (1900-1958). Verificar que elas têm valores próprios 1 e -1. Determinar depois todas as matrizes 2×2 com elementos complexos tendo os dois valores próprios 1 e -1.

5. Determinar todas as matrizes 2×2 com elementos reais cujos valores próprios são (a) reais e distintos, (b) reais e iguais, (c) complexos conjugados.
6. Determinar a, b, c, d, e, f , dados que os vectores $(1, 1, 1)$, $(1, 0, -1)$ e $(1, -1, 0)$ são vectores próprios da matriz

$$\begin{bmatrix} 1 & 1 & 1 \\ a & b & c \\ d & e & f \end{bmatrix}$$

7. Calcular os valores próprios e vectores próprios de cada uma das matrizes seguintes e calcular igualmente a dimensão do espaço $E(\lambda)$ para cada valor próprio λ .

(a) $\begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 4 & -7 & 1 \end{bmatrix}$, (b) $\begin{bmatrix} 2 & 1 & 3 \\ 1 & 2 & 3 \\ 3 & 3 & 20 \end{bmatrix}$, (c) $\begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix}$.

8. Calcular os valores próprios de cada uma das cinco matrizes.

(a) $\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$, (b) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$, (c) $\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$,

(d) $\begin{bmatrix} 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \end{bmatrix}$, (e) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$.

Estas chamam-se *matrizes de Dirac* em homenagem a Paul A.M. Dirac (1902-), físico inglês. Elas aparecem na resolução da equação das ondas relativista em Mecânica Quântica.

9. Se A e B são matrizes $n \times n$, com B uma matriz diagonal, provar (por indução) que o determinante $f(\lambda) = \det(\lambda B - A)$ é um polinómio em λ com $f(0) = (-1)^n \det A$, e com o coeficiente de λ^n igual ao produto dos elementos principais de B .
10. Provar que uma matriz quadrada A e a sua transposta A' têm o mesmo polinómio característico.
11. Se A e B são matrizes $n \times n$, com A não singular, provar que AB e BA têm o mesmo conjunto de valores próprios.

Nota: Pode demonstrar-se que AB e BA têm o mesmo polinómio característico, mesmo se A é singular, mas não se pede a demonstração para esta hipótese.

12. Seja A uma matriz $n \times n$ com o polinómio característico $f(\lambda)$. Provar (por indução) que o coeficiente de λ^{n-1} em $f(\lambda)$ é $-\text{tr } A$.
13. Sejam A e B matrizes $n \times n$ com $\det A = \det B$ e $\text{tr } A = \text{tr } B$. Provar que A e B têm o mesmo polinómio característico se $n = 2$, mas que tal não se verifica se $n > 2$.
14. Provar cada uma das seguintes proposições relativas ao traço.

$$(a) \text{tr } (A + B) = \text{tr } A + \text{tr } B.$$

$$(b) \text{tr } (cA) = c \text{tr } A.$$

$$(c) \text{tr } (AB) = \text{tr } (BA).$$

$$(d) \text{tr } A^t = \text{tr } A.$$

4.9. Matrizes representando a mesma transformação linear. Matrizes semelhantes

Nesta Secção vamos demonstrar que duas representações matriciais distintas de uma transformação linear têm o mesmo polinómio característico. Para isso vamos analisar mais em pormenor a relação entre matrizes que representem a mesma transformação.

Recordemos como se definem as representações matriciais. Admitamos que $T: V \rightarrow W$ é uma aplicação de um espaço n dimensional V num espaço m dimensional W . Sejam (e_1, \dots, e_n) e (W_1, \dots, W_m) bases ordenadas para V e W respectivamente. A representação matricial de T relativa a esta escolha de bases é a matriz $m \times n$ cujas colunas são as componentes de $T(e_1), \dots, T(e_n)$ relativamente à base (w_1, \dots, w_m) . Diferentes representações matriciais são devidas a diferentes escolhas das bases.

Consideremos agora o caso em que $V = W$, e admitamos que se considera a mesma base ordenada (e_1, \dots, e_n) em V e W . Seja $A = (a_{ik})$ a matriz de T relativa a esta base. Isto significa que se tem

$$(4.8) \quad T(e_k) = \sum_{i=1}^n a_{ik} e_i \quad \text{para } k = 1, 2, \dots, n.$$

Escolhamos agora outra base ordenada (u_1, \dots, u_n) para V e W e seja $B = (b_{kj})$ a matriz de T relativa a esta nova base. Então tem-se

$$(4.9) \quad T(u_j) = \sum_{k=1}^n b_{kj} u_k \quad \text{para } j = 1, 2, \dots, n.$$

Visto que cada u_j pertence ao espaço gerado por e_1, \dots, e_n podemos escrever

$$(4.10) \quad u_j = \sum_{k=1}^n c_{kj} e_k \quad \text{para } j = 1, 2, \dots, n,$$

para algum conjunto de escalares c_{kj} . Então a matriz $C = (c_{kj})$, $n \times n$, definida por estes escalares é não singular porque representa uma transformação linear que aplica uma base de V sobre outra base de V . Aplicando T a ambos os membros de (4.10) resultam igualmente as equações

$$(4.11) \quad T(u_j) = \sum_{k=1}^n c_{kj} T(e_k) \quad \text{para } j = 1, 2, \dots, n.$$

Os sistemas de equações de (4.8) a (4.11) podem escrever-se de modo mais simples na forma matricial introduzindo matrizes cujos elementos sejam vectores. Sejam

$$E = [e_1, \dots, e_n] \quad \text{e} \quad U = [u_1, \dots, u_n]$$

matrizes linha, $1 \times n$ cujos elementos são os das bases que se consideram. Então o conjunto de equações em (4.10) pode escrever-se como uma simples equação matricial

$$(4.12) \quad U = EC.$$

Analogamente, se introduzimos

$$E' = [T(e_1), \dots, T(e_n)] \quad \text{e} \quad U' = [T(u_1), \dots, T(u_n)],$$

as equações (4.8), (4.9) e (4.11) escrevem-se, respectivamente,

$$(4.13) \quad E' = EA, \quad U' = UB, \quad U' = E'C.$$

De (4.12) resulta

$$E = UC^{-1}.$$

Para encontrarmos a relação entre A e B exprimimos U' de duas maneiras diferentes em função de U . De (4.13) resulta

$$U' = UB$$

e

$$U' = E'C = EAC = UC^{-1}AC.$$

Portanto $UB = UC^{-1}AC$. Mas cada elemento nesta equação matricial é uma combinação linear dos vectores base u_1, \dots, u_n . Porque os u_i são independentes teremos

$$B = C^{-1}AC.$$

Demonstrámos assim o seguinte:

TEOREMA 4.6. *Se duas matrizes $n \times n$, A e B , representam a mesma transformação linear T , então existe uma matriz não singular C tal que*

$$B = C^{-1}AC.$$

Além disso, se A é a matriz de T relativa à base $E = [e_1, \dots, e_n]$ é se B é a matriz de T relativa à base $U = [u_1, \dots, u_n]$, então para C pode tomar-se a matriz não singular que relaciona as duas bases através da equação matricial $U = EC$.

O inverso do teorema 4.6 é também verdadeiro.

TEOREMA 4.7. Se A e B são duas matrizes $n \times n$ relacionadas pela equação da forma $B = C^{-1}AC$, com C uma matriz $n \times n$ não singular, então A e B representam a mesma transformação linear.

Demonstração. Escolhamos uma base $E = [e_1, \dots, e_n]$ para um espaço n -dimensional V . Sejam u_1, \dots, u_n vectores definidos por

$$(4.14) \quad u_j = \sum_{k=1}^n c_{kj} e_k \quad \text{para } j = 1, 2, \dots, n,$$

onde os escalares c_{kj} são os elementos de C . Visto C ser não singular representa uma transformação linear invertível, pelo que $U = [u_1, \dots, u_n]$ é também uma base para V e tem-se $U = EC$.

Seja T a transformação linear que admite A como representação matricial relativa à base E , e seja S a transformação que admite B como representação relativa à base U . Tem-se então

$$(4.15) \quad T(e_k) = \sum_{i=1}^n a_{ik} e_i \quad \text{para } k = 1, 2, \dots, n$$

e

$$(4.16) \quad S(u_j) = \sum_{k=1}^n b_{kj} u_k \quad \text{para } j = 1, 2, \dots, n.$$

Demonstraremos que $S = T$ provando que $T(u_j) = S(u_j)$ para cada j .

As equações (4.15) e (4.16) podem escrever-se, de maneira simplificada, na forma matricial.

$$[T(e_1), \dots, T(e_n)] = EA, \quad [S(u_1), \dots, S(u_n)] = UB.$$

Aplicando T a (4.14) obtemos também a relação $T(u_j) = \sum c_{kj} T(e_k)$ ou

$$[T(u_1), \dots, T(u_n)] = EAC.$$

Mas sabe-se que

$$UB = ECB = EC(C^{-1}AC) = EAC,$$

o que prova que $T(u_j) = S(u_j)$ para cada j . Deste modo $T(x) = S(x)$ para todo x em V , pelo que $T = S$. Por outras palavras, as matrizes A e B representam a mesma transformação linear.

DEFINIÇÃO. Duas matrizes $n \times n$, A e B , dizem-se semelhantes se existir uma matriz não singular C tal que $B = C^{-1}AC$.

Os teoremas 4.6 e 4.7 podem combinar-se para dar origem a:

TEOREMA 4.8. Duas matrizes $n \times n$ são semelhantes se e só se representam a mesma transformação linear.

Matrizes semelhantes gozam de muitas propriedades importantes. Por exemplo, têm o mesmo determinante visto que

$$\det(C^{-1}AC) = \det(C^{-1})(\det A)(\det C) = \det A.$$

Esta propriedade conduz-nos ao seguinte teorema.

TEOREMA 4.9. Matrizes semelhantes tem o mesmo polinómio característico e portanto os mesmos valores próprios.

Demonstração. Se A e B são semelhantes, existe uma matriz não singular C tal que $B = C^{-1}AC$. Portanto temos

$$\lambda I - B = \lambda I - C^{-1}AC = \lambda C^{-1}IC - C^{-1}AC = C^{-1}(\lambda I - A)C.$$

Isto mostra que $\lambda I - B$ e $\lambda I - A$ são semelhantes, pelo que $\det(\lambda I - B) = \det(\lambda I - A)$.

Os teoremas 4.8 e 4.9 em conjunto mostram que todas as representações matriciais de uma dada transformação linear T têm o mesmo polinómio característico. Este polinómio chama-se também o polinómio característico de T .

O teorema que se segue é uma combinação dos teoremas 4.5, 4.2 e 4.6. No teorema 4.10, F representa quer o corpo dos reais \mathbf{R} quer o corpo complexo \mathbf{C} .

TEOREMA 4.10. Se $T: V \rightarrow V$ é uma transformação linear com F o corpo dos escalares de V e $\dim V = n$, e se se admite que o polinómio característico de T tem n raízes distintas $\lambda_1, \dots, \lambda_n$ em F , então:

- Os correspondentes vectores próprios u_1, \dots, u_n forma uma base para V .
- A matriz de T relativa à base ordenada $U = \{u_1, \dots, u_n\}$ é a matriz diagonal Λ tendo os valores próprios como elementos principais:

$$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n).$$

(c) Se A é a matriz de T relativa a outra base $E = \{e_1, \dots, e_n\}$ então.

$$\Lambda = C^{-1}AC,$$

onde C é a matriz não singular relacionando as duas bases pela equação

$$U = EC.$$

Demonstração. Pelo teorema 4.5 cada raiz λ_i é um valor próprio. Porque existem n raízes distintas, o teorema 4.2 diz-nos que os correspondentes vectores próprios u_1, \dots, u_n são independentes. Por conseguinte formam uma base para V , e a alínea (a) fica demonstrada. Visto que $T(u_i) = \lambda_i u_i$ a matriz de T relativa a U é a matriz diagonal Λ , o que demonstra (b). Para provar (c) usamos o teorema 4.6.

Nota: A matriz não singular C do teorema 4.10 chama-se uma *matriz diagonalizadora*. Se (e_1, \dots, e_n) é uma base dos vectores coordenados unitários (I_1, \dots, I_n) então a equação $U = EC$ no teorema 4.10 mostra que a coluna de ordem k de C é formada pelas componentes do vector próprio u_k relativo a (I_1, \dots, I_n) .

Se os valores próprios de A são distintos, então A é semelhante a uma matriz diagonal. Se os valores próprios não são distintos, então A ainda pode ser semelhante a uma matriz diagonal. Isto verifica-se-á se e só se existirem k vectores próprios independentes correspondentes a cada valor próprio de multiplicidade k . Nos exercícios que se seguem são dados alguns exemplos.

4.10. Exercícios

1. Provar que as matrizes $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ e $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ têm os mesmos valores próprios, mas não são semelhantes.
2. Para cada alínea determinar uma matriz não singular C tal que $C^{-1}AC$ seja uma matriz diagonal ou explicar porque razão não existe tal C .

$$(a) A = \begin{bmatrix} 1 & 0 \\ 1 & 3 \end{bmatrix}, \quad (b) A = \begin{bmatrix} 1 & 2 \\ 5 & 4 \end{bmatrix}, \quad (c) A = \begin{bmatrix} 2 & 1 \\ -1 & 4 \end{bmatrix}, \quad (d) A = \begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix}.$$

3. Dão-se três bases no plano. Com respeito a estas bases um ponto tem coordenadas (x_1, x_2) , (y_1, y_2) e (z_1, z_2) respectivamente. Admitindo que $[y_1, y_2] = [x_1, x_2]A$, $[z_1, z_2] = [x_1, x_2]B$, e $[z_1, z_2] = [y_1, y_2]C$, com A , B e C matrizes 2×2 , exprimir C em função de A e B .
4. Para cada alínea, mostrar que os valores próprios de A não são distintos mas que A tem três vectores próprios independentes. Determinar uma matriz não singular C tal que $C^{-1}AC$ seja uma matriz diagonal.

$$(a) A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad (b) A = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 3 & 1 \\ -1 & -1 & 1 \end{bmatrix}.$$

5. Mostra que nenhuma das seguintes matrizes é semelhante a uma matriz diagonal, mas que cada uma é semelhante a uma matriz triangular da forma $\begin{bmatrix} \lambda & 0 \\ 1 & \lambda \end{bmatrix}$, com λ um valor próprio.

$$(a) \begin{bmatrix} 2 & -1 \\ 0 & 2 \end{bmatrix}, \quad (b) \begin{bmatrix} 2 & 1 \\ -1 & 4 \end{bmatrix}.$$

6. Determinar os valores próprios e vectores próprios da matriz $\begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 1 \\ -1 & -3 & 3 \end{bmatrix}$ e com eles provar que não é semelhante a uma matriz diagonal.

7. (a) Provar que uma matriz quadrada A é não singular se e só se 0 não é um valor próprio de A .
 (b) Se A é não singular, provar que os valores próprios de A^{-1} são os recíprocos dos valores próprios de A .
8. Dada uma matriz $n \times n$, A , com elementos reais e tal que $A^2 = -I$, provar que as seguintes proposições relativas a A .
- (a) A é não singular.
 - (b) n é par.
 - (c) A não tem valores próprios reais.
 - (d) $\det A = 1$.

VALORES PRÓPRIOS DE OPERADORES EM ESPAÇOS EUCLIDIANOS

5.1. Valores próprios e produtos internos

Este capítulo descreve algumas propriedades dos valores próprios e vectores próprios de transformações lineares que operam em espaços euclidianos, isto é, espaços lineares dotados com a estrutura de produto interno. Recordemos as propriedades fundamentais do produto interno.

Num espaço euclidiano *real* o produto interno (x, y) de dois elementos x e y é um número real satisfazendo às seguintes propriedades:

- (1) $(x, y) = (y, x)$ (simetria)
- (2) $(x + z, y) = (x, y) + (z, y)$ (linearidade)
- (3) $(cx, y) = c(x, y)$ (homogeneidade)
- (4) $(x, x) > 0$ se $x \neq 0$ (positividade)

Num espaço euclidiano *complexo* o produto interno é um número complexo satisfazendo às mesmas propriedades, com excepção da simetria que é substituída pela *simetria hermitica*.

$$(1') \quad (x, y) = \overline{(y, x)},$$

onde a barra representa o complexo conjugado. Em (3) o escalar c é complexo. De (1') e (3) resulta

$$(3') \quad (x, cy) = \bar{c}(x, y),$$

que significa que os escalares são os conjugados quando são tomados do segundo factor. Tomando $x = y$ em (1') vemos que (x, y) é real, pelo que a propriedade (4) tem significado se o espaço é complexo.

Quando usamos a expressão espaço euclidiano sem qualquer outra designação, deve entender-se que o espaço pode ser real ou complexo. Embora a maior parte das

nossas aplicações digam respeito a espaços de dimensão finita, não se impõe esta restrição à priori.

O primeiro teorema vai mostrar que os valores próprios (se existirem) poderão exprimir-se em função do produto interno.

TEOREMA 5.1. *Se E representa um espaço euclidiano, V um subespaço de E , e $T: V \rightarrow E$ uma transformação linear admitindo um valor próprio λ com um correspondente vector próprio x , então*

$$(5.1) \quad \lambda = \frac{(T(x), x)}{(x, x)}.$$

Demonstração. Uma vez que $T(x) = \lambda x$ tem-se

$$(T(x), x) = (\lambda x, x) = \lambda(x, x).$$

Porque $x \neq 0$, podemos dividir por (x, x) para obtermos (5.1).

Várias propriedades dos valores próprios podem facilmente deduzir-se das equações (5.1). Por exemplo, da simetria hermitica do produto interno temos a fórmula análoga

$$(5.2) \quad \bar{\lambda} = \frac{(x, T(x))}{(x, x)}$$

para o complexo conjugado de λ . De (5.1) e (5.2) vemos que λ é real ($\lambda = \bar{\lambda}$) se e só se $(T(x), x)$ é real, isto é, se e só se

$$(T(x), x) = (x, T(x)) \quad \text{para o vector próprio } x.$$

(Esta condição é trivial num espaço euclidiano real). Também λ é imaginário puro ($\lambda = \bar{\lambda}$) se e só se $(T(x), x)$ é imaginário puro, isto é, se e só se

$$(T(x), x) = -(x, T(x)) \quad \text{para o vector próprio } x.$$

5.2. Transformações hermiticas e hemi-hermiticas

Nesta Secção vamos fazer a introdução de dois tipos importantes de operadores lineares definidos para espaços euclidianos. Estes operadores têm duas espécies de designação, que dependem do facto do espaço euclidiano fundamental ter um produto interno real ou complexo. No caso real as transformações dizem-se *simétricas* e *hemi-simétricas*. No caso chamam-se *hermiticas* e *hemi-hermiticas*. Estas transformações aparecem em enorme variedade de aplicações distintas. Por exemplo, operadores hermiticos definidos em espaços de dimensão infinita desempenham um papel importante na mecânica quântica. Discutiremos fundamentalmente o caso complexo visto que não apresenta dificuldades suplementares.

DEFINIÇÃO. Seja E um espaço euclidiano e V um subespaço de E . Uma transformação linear $T: V \rightarrow E$ diz-se *hermítica em V* se

$$(T(x), y) = (x, T(y)) \quad \text{para todo o par } x, y \text{ em } V.$$

O operador T diz-se *hemi-hermítico em V* se

$$(T(x), y) = -(x, T(y)) \quad \text{para todo o par } x, y \text{ em } V.$$

Por outras palavras, um operador hermítico T pode passar-se de um para outro factor de um produto interno sem que o valor deste se altere. Efectuando essa mesma troca com um operador hemi-hermítico o produto interno muda de sinal.

Nota: Como já mencionámos, se E é um espaço euclidiano *real*, as transformações hermíticas dizem-se *simétricas*; transformações *hemi-hermíticas* dizem-se *hemi-simétricas*.

EXEMPLO 1. *Simetria e hemi-simetria no espaço $C(a, b)$.* Seja $C(a, b)$ o espaço de todas as funções reais contínuas num intervalo fechado $[a, b]$, com o produto interno real

$$(f, g) = \int_a^b f(t)g(t) dt.$$

Seja V um subespaço de $C(a, b)$. Se $T: V \rightarrow C(a, b)$ é uma transformação linear então $(f, T(g)) = \int_a^b f(t)Tg(t)dt$, onde $Tg(t)$ significa $T(g)(t)$. Deste modo as condições para simetria e hemi-simetria vêm

$$(5.3) \quad \int_a^b \{f(t)Tg(t) - g(t)Tf(t)\} dt = 0 \quad \text{se } T \text{ é simétrico,}$$

$$(5.4) \quad \int_a^b \{f(t)Tg(t) + g(t)Tf(t)\} dt = 0 \quad \text{se } T \text{ é hemi-simétrico.}$$

EXEMPLO 2. *Multiplicação por uma função dada.* No espaço $C(a, b)$ do Exemplo 1, escolhamos uma função p e definamos $T(f) = pf$, o produto de p por f . Para este T , a equação (5.3) é satisfeita quaisquer que sejam f e g em $C(a, b)$ visto que a função integranda é zero. Por conseguinte, a multiplicação por uma função fixa é um operador simétrico.

EXEMPLO 3. *O operador derivação.* No espaço $C(a, b)$ do Exemplo 1, seja V o subespaço de todas as funções f que admitem derivada contínua no intervalo aberto (a, b) e que verificam ainda as condições de fronteira $f(a) = f(b)$. Seja $D: V \rightarrow C(a, b)$ o operador derivação dado por $D(f) = f'$. É fácil provar que D é hemi-simétrico. Neste caso a função integranda (5.4) é a derivada do produto fg , pelo que o integral é igual a

$$\int_a^b (fg)'(t) dt = f(b)g(b) - f(a)g(a).$$

Uma vez que ambos f e g satisfazem as condições limites, temos $f(b)g(b) - f(a)g(a) = 0$. Assim, as condições limites implicam a hemi-simetria de D . As únicas funções próprias no subespaço V são as funções constantes. Elas dizem respeito ao valor próprio 0.

EXEMPLO 4. Operadores de Sturm-Liouville. Este exemplo é importante na teoria das equações diferenciais lineares de segunda ordem. Consideremos os espaços $C(a, b)$ do Exemplo 1 uma vez mais, e seja V o subespaço de todas as funções f que admitem derivada de segunda ordem contínua em $[a, b]$ e que satisfazem igualmente às duas condições de fronteira

$$(5.5) \quad p(a)f(a) = 0, \quad p(b)f(b) = 0,$$

onde p é uma função de $C(a, b)$ com derivada contínua em $[a, b]$. Seja q outra função dada em $C(a, b)$ e $T: V \rightarrow C(a, b)$ o operador definido pela equação

$$T(f) = (pf')' + qf.$$

Este é o chamado *operador de Sturm-Liouville*. Para averiguar da sua simetria observamos que $fT(g) - gT(f) = f(pg') - g(pf')$. Utilizando esta igualdade em (5.3) e integrando ambos $\int_a^b f \cdot (pg')' dt$ e $\int_a^b g \cdot (pf')' dt$ por partes, encontramos

$$\int_a^b \{fT(g) - gT(f)\} dt = fpg' \Big|_a^b - \int_a^b pg'f' dt - gpf' \Big|_a^b + \int_a^b pf'g' dt = 0,$$

uma vez que f e g satisfazem ambas às condições (5.5). Por conseguinte T é simétrico em V . As funções próprias de T são aquelas funções não nulas f que satisfazem, para algum real λ , à equação diferencial da forma

$$(pf')' + qf = \lambda f$$

em $[a, b]$, e também satisfazem às condições (5.5).

5.3. Valores próprios e vectores próprios de operadores hermiticos e hemi-hermiticos

Com respeito aos valores próprios temos o seguinte:

TEOREMA 5.2. Se T admite um valor próprio λ , então:

(a) Se T é hermitico, λ é real: $\lambda = \bar{\lambda}$.

(b) Se T é hemi-hermitico, λ é imaginário puro: $\lambda = -\bar{\lambda}$.

Demonstração. Seja x um vector próprio correspondente a λ . Então temos

$$\lambda = \frac{(T(x), x)}{(x, x)} \quad \text{e} \quad \bar{\lambda} = \frac{(x, T(x))}{(x, x)}.$$

Se T é hermitico temos $(T(x), x) = (x, T(x))$ pelo que $\lambda = \bar{\lambda}$. Se T é hemi-hermitico temos $(T(x), x) = -(x, T(x))$ pelo que $\lambda = -\bar{\lambda}$.

Nota: Se T é *simétrico*, o teorema 5.2 não nos diz nada acerca dos valores próprios de T , uma vez que todos os valores próprios devem ser reais se o produto interno é real. Se T é *hemi-simétrico*, os valores próprios de T devem ser simultaneamente reais e imaginários puros e por conseguinte todos os valores próprios de operadores hemi-simétricos devem ser zeros (caso existam).

5.4. Ortogonalidade de vectores próprios correspondentes a valores próprios distintos

Valores próprios distintos de uma transformação linear qualquer correspondem a vectores próprios independentes (pelo teorema 4.2). Para as transformações hermiticas e hemi-hermiticas podemos afirmar algo mais.

TEOREMA 5.3. *Se T é uma transformação hermitica ou hemi-hermitica e λ e μ são valores próprios distintos de T com correspondentes vectores próprios x e y , então x e y são ortogonais, isto é, $(x, y) = 0$.*

Demonstração. Escrevemos $T(x) = \lambda x$, $T(y) = \mu y$ e comparemos os dois produtos internos $(T(x), y)$ e $(x, T(y))$. Temos

$$(T(x), y) = (\lambda x, y) = \lambda(x, y) \quad \text{e} \quad (x, T(y)) = (x, \mu y) = \bar{\mu}(x, y).$$

Se T é hermitico podemos escrever $\lambda(x, y) = \bar{\mu}(x, y) = \mu(x, y)$ visto que $\bar{\mu} = \mu$. Deste modo $(x, y) = 0$ por ser $\lambda \neq \mu$. Se T é hemi-hermitico obtemos $\lambda(x, y) = -\bar{\mu}(x, y) = -\mu(x, y)$ o que implica $(x, y) = 0$.

EXEMPLO. Apliquemos o teorema 5.3 às funções não nulas que satisfazem à equação diferencial da forma

$$(5.6) \quad (pf')' + qf = \lambda f$$

num intervalo $[a, b]$ e as quais verificam também as condições $p(a)f(a) = p(b)f(b) = 0$. A conclusão é que duas quaisquer soluções f e g correspondendo a dois valores distintos de λ são ortogonais. Por exemplo, consideremos a equação diferencial do movimento harmónico simples

$$f'' + k^2 f = 0$$

no intervalo $[0, \pi]$, com $k \neq 0$. Esta é a forma de (5.6) com $p = 1$, $q = 0$ e $\lambda = -k^2$. Todas as soluções são dadas por $f(t) = c_1 \cos kt + c_2 \sin kt$. A condição de fronteira $f(0) = 0$ implica $c_1 = 0$. A segunda condição, $f(\pi) = 0$, implica $c_2 \sin k\pi = 0$. Porque $c_2 \neq 0$ para

uma solução não nula, deve ter-se $\sin k\pi = 0$, o que significa que k é um inteiro. Por outras palavras, soluções não nulas que verifiquem as condições de fronteira existem se e só se k for inteiro. Estas soluções são $f(t) = \sin nt$, $n = \pm 1, \pm 2, \dots$. A condição de ortogonalidade implicada pelo teorema 5.3 transforma-se aqui na relação conhecida

$$\int_0^\pi \sin nt \sin mt \, dt = 0$$

se m^2 e n^2 são inteiros distintos.

5.5. Exercícios

1. Seja E um espaço euclidiano, V um subespaço e $T: V \rightarrow E$ uma dada transformação linear. Seja λ um escalar e x um elemento não nulo de V . Provar que λ é um valor próprio de T com x como vector próprio se e só se

$$(T(x), y) = \lambda(x, y) \quad \text{para todo } y \text{ em } E.$$

2. Seja $T(x) = cx$ para todo x um espaço linear V , sendo c um escalar determinado. Provar que T é simétrico se V é um espaço euclidiano real.
3. Suponhamos $T: V \rightarrow V$ uma transformação hermitica.
 - (a) Provar que T^n é hermitica para todo inteiro positivo n , e que T^{-1} é hermitico se T é invertível.
 - (b) O que pode concluir-se acerca de T^n e T^{-1} se T é hemi-hermitico?
4. Sejam $T_1: V \rightarrow E$ e $T_2: V \rightarrow E$ duas transformações hermiticas.
 - (a) Provar que $aT_1 + bT_2$ é hermitica para todo o par de escalares reais a e b .
 - (b) Provar que o produto (composição) T_1T_2 é hermitica se T_1 e T_2 comutam, isto é, se $T_1T_2 = T_2T_1$.
5. Seja $V = V_2(\mathbb{R})$ com o produto escalar como produto interno. Seja T uma simetria no plano XOY , isto é, seja $T(i) = i$, $T(j) = j$ e $T(k) = -k$. Provar que T é simétrico.
6. Seja $C(0, 1)$ o espaço linear real de todas as funções reais contínuas em $[0, 1]$ com o produto interno $(f, g) = \int_0^1 f(t)g(t)dt$. Seja V o subespaço de todos os f tais que $\int_0^1 f(t)dt = 0$. Seja $T: V \rightarrow C(0, 1)$ o operador integração definido por $Tf(x) = \int_0^x f(t)dt$. Provar que T é hemi-simétrico.
7. Seja V o espaço euclidiano real de todos os polinómios com o produto interno $(f, g) = \int_{-1}^1 f(t)g(t)dt$. Determinar quais das seguintes transformações $T: V \rightarrow V$ são simétricas ou hemi-simétricas.

(a) $Tf(x) = f(-x)$.	(c) $Tf(x) = f(x) + f(-x)$.
(b) $Tf(x) = f(x)f(-x)$.	(d) $Tf(x) = f(x) - f(-x)$.
8. No Exemplo 4 da Secção 5.2 modificar o produto interno do modo seguinte:

$$(f, g) = \int_a^b f(t)g(t)w(t) \, dt,$$

onde w é uma determinada função positiva em $C(a, b)$. Modificar o operador de Sturm-Liouville T escrevendo

$$T(f) = \frac{(pf')' + qf}{w}.$$

Provar que este operador modificado é simétrico no subespaço V .

9. Seja V um subespaço de um espaço euclidiano complexo E . Seja $T: V \rightarrow E$ uma transformação linear e defina-se uma função de valores escalares Q em V do modo seguinte:

$$Q(x) = (T(x), x) \quad \text{para qualquer } x \text{ em } V.$$

- Se T é hermitica em V , provar que $Q(x)$ é real para todo x .
 - Se T é hemi-hermitica, provar que $Q(x)$ é imaginária pura para todo x .
 - Provar que $Q(tx) = tQ(x)$ para qualquer escalar t .
 - Provar que $Q(x+y) = Q(x) + Q(y) + (T(x), y) + (T(y), x)$, e determinar a fórmula correspondente para $Q(x+ty)$.
 - Se $Q(x) = 0$ para todo x provar que $T(x) = 0$ para todo x .
 - Se $Q(x)$ é real para todo x provar que T é hermitica. [Sugestão: Utilizar o facto de que $Q(x+ty)$ é igual à sua conjugada para todo o escalar t .]
10. Este exercício põe em evidência que os polinómios de Legendre (introduzidos na Secção 1.14) são funções próprias do operador de Sturm-Liouville. Os polinómios de Legendre são definidos pela equação

$$P_n(t) = \frac{1}{2^n n!} f_n^{(n)}(t), \quad \text{onde} \quad f_n(t) = (t^2 - 1)^n.$$

- Verificar que $(t^2 - 1)f_n'(t) = 2nt f_n(t)$.
- Derivar $n+1$ vezes a equação (a), mediante a fórmula de Leibniz, (ver p. 260 do volume I) para obter

$$(t^2 - 1)f_n^{(n+2)}(t) + 2t(n+1)f_n^{(n+1)}(t) + n(n+1)f_n^{(n)}(t) = 2ntf_n^{(n+1)}(t) + 2n(n+1)f_n^{(n)}(t).$$

- Mostrar que a equação em (b) pode ser de novo escrita na forma

$$[(t^2 - 1)P_n'(t)]' = n(n+1)P_n(t).$$

Isto mostra que $P_n(t)$ é uma função própria do operador de Sturm-Liouville T dado no intervalo $[-1, 1]$ por $T(f) = (pf')'$, onde $p(t) = t^2 - 1$. A função própria $P_n(t)$ corresponde ao valor próprio $\lambda = n(n+1)$. Neste exemplo as condições de fronteira para a simetria são automaticamente satisfeitos visto $p(1) = p(-1) = 0$.

5.6. Existência de um conjunto ortonormal de vectores próprios para operadores hermiticos e hemi-hermiticos em espaços de dimensão finita

Ambos os teoremas 5.2 e 5.3 se baseiam na afirmação de que T admite um valor próprio. Como sabemos os valores próprios não existem necessariamente. Contudo, se T actua num espaço complexo de dimensão finita, então os valores próprios existem sempre uma vez que são as raízes do polinómio característico. Se T é hermitico, todos

os valores próprios são reais. Se T é hemi-hermítico, todos os valores próprios são imaginários puros.

Sabemos também que dois valores próprios distintos dizem respeito a vectores próprios ortogonais, se T é hermitico ou hermi-hermítico. Servindo-nos desta propriedade podemos provar que T admite um conjunto ortonormado de vectores próprios que geram todo o espaço. (Lembramos que um conjunto ortogonal se diz ortonormado se cada um dos seus elementos tem norma 1).

TEOREMA 5.4. *Se $\dim V = n$ e $T: V \rightarrow V$ é hermitico ou hemi-hermítico, então existem n vectores próprios u_1, \dots, u_n que constituem uma base ortonormada para V . Consequentemente a matriz de T relativa a esta base é a matriz diagonal $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, com λ_k o valor próprio respeitante a u_k .*

Demonstração: Recorremos à demonstração por indução em n . Se $n = 1$, então T tem precisamente um valor próprio. Qualquer vector próprio u_1 de norma 1 é uma base ortonormada para V .

Suponhamos que o teorema é verdadeiro para todo o espaço euclidiano de dimensão $n - 1$. Para demonstrar que é também verdadeiro para V escolhamos um valor próprio λ_1 para T e um vector próprio correspondente u_1 de norma 1. Então $T(u_1) = \lambda_1 u_1$ e $\|u_1\| = 1$. Seja S o subespaço gerado por u_1 . Devemos aplicar a hipótese de indução ao subespaço S^\perp formado por todos os elementos de V que são ortogonais a u_1 ,

$$S^\perp = \{x \mid x \in V, (x, u_1) = 0\}.$$

Para tal necessitamos saber que $\dim S^\perp = n - 1$ e que T aplica S^\perp em si próprio.

Pelo teorema 1.7(a) sabemos que u_1 é parte de uma base de V , por exemplo a base (u_1, v_2, \dots, v_n) . Podemos supôr, sem perda de generalidade, que esta é uma base ortonormada. (Caso contrário aplicamos o método de Gram-Schmidt para a transformar numa base ortonormada, mantendo u_1 como primeiro elemento da base). Tomemos agora qualquer x em S^\perp e escrevamos

$$x = x_1 u_1 + x_2 v_2 + \dots + x_n v_n.$$

Então $x_1 = (x, u_1) = 0$ visto que a base é ortonormada, pelo que x pertence ao espaço gerado por v_2, \dots, v_n . Consequentemente $\dim S^\perp = n - 1$.

Demonstremos em seguida que T aplica S^\perp em si próprio. Suponhamos que T é hermitico. Se $x \in S^\perp$ temos

$$(T(x), u_1) = (x, T(u_1)) = (x, \lambda_1 u_1) = \lambda_1 (x, u_1) = 0,$$

pelo que $T(x) \in S^\perp$. Uma vez que T é hermitico em S^\perp podemos aplicar a hipótese de indução para concluir que T admite $n - 1$ vectores próprios u_2, \dots, u_n os quais formam uma base ortonormada para S^\perp . Deste modo o conjunto ortogonal u_1, \dots, u_n é uma

base ortonormada para V , e o teorema fica demonstrado na hipótese de T ser hermitico. Um raciocinio semelhante é válido no caso de T ser hemi-hermitico.

5.7. Representação matricial de operadores hermiticos e hemi-hermiticos

Nesta Secção supomos que V é um espaço euclidiano de dimensão finita. Uma transformação hermitica ou hemi-hermitica pode caracterizar-se pela sua acção sobre os elementos de uma base qualquer.

TEOREMA 5.5. *Se (e_1, \dots, e_n) é uma base de V e $T: V \rightarrow V$ uma transformação linear, então:*

- (a) *T é hermitica se e só se $(T(e_j), e_i) = (e_j, T(e_i))$ para quaisquer i e j .*
- (b) *T é hemi-hermitica se e só se $(T(e_j), e_i) = -(e_j, T(e_i))$ para quaisquer i e j .*

Demonstração. Consideremos dois quaisquer elementos x e y de V e exprimamos cada um deles como combinação linear dos elementos da base, seja $x = \sum x_j e_j$ e $y = \sum y_i e_i$. Então temos

$$(T(x), y) = \left(\sum_{j=1}^n x_j T(e_j), y \right) = \sum_{j=1}^n x_j \left(T(e_j), \sum_{i=1}^n y_i e_i \right) = \sum_{j=1}^n \sum_{i=1}^n x_j y_i (T(e_j), e_i).$$

Analogamente encontramos

$$(x, T(y)) = \sum_{j=1}^n \sum_{i=1}^n x_j y_i (e_j, T(e_i)).$$

As proposições (a) e (b) são consequência imediata destas equações.

Caracterizemos agora estes conceitos por intermédio da representação matricial de T .

TEOREMA 5.6. *Se (e_1, \dots, e_n) for uma base ortonormada para V e $A = (a_{ij})$ a representação matricial de uma transformação linear $T: V \rightarrow V$ relativa a esta base então tem-se que:*

- (a) *T é hermitica se e só se $a_{ij} = \bar{a}_{ji}$ para quaisquer i e j .*
- (b) *T é hemi-hermitica se e só se $a_{ij} = -\bar{a}_{ji}$ para quaisquer i e j .*

Demonstração. Uma vez que A é a matriz de T temos $T(e_j) = \sum_{k=1}^n a_{kj} e_k$. Tomando o produto interno de $T(e_j)$ por e_i e recorrendo à linearidade deste produto obtemos

$$(T(e_j), e_i) = \left(\sum_{k=1}^n a_{kj} e_k, e_i \right) = \sum_{k=1}^n a_{kj} (e_k, e_i).$$

Mas $(e_k, e_i) = 0$, salvo se $k = i$, pelo que a última soma se simplifica e se reduz a $a_{ij}(e_i, e_i)$, visto que $(e_i, e_i) = 1$. Consequentemente temos

$$a_{ij} = (T(e_j), e_i) \quad \text{para quaisquer } i \text{ e } j.$$

Trocando entre si os índices i e j , tomando os conjugados, e utilizando a simetria hermitica do produto interno, encontramos

$$\bar{a}_{ji} = (e_j, T(e_i)) \quad \text{para quaisquer } i \text{ e } j.$$

Para completar a demonstração recorreremos ao teorema 5.5.

5.8. Matrizes hermiticas e hemi-hermiticas. A associada de uma matriz

O teorema 5.6 sugere a seguinte definição.

DEFINIÇÃO. Uma matriz quadrada $A = (a_{ij})$ diz-se *hermitica* se $a_{ij} = \bar{a}_{ji}$ para i e j quaisquer. A matriz diz-se *hemi-hermitica* se $a_{ij} = -\bar{a}_{ji}$ para i e j quaisquer.

O teorema 5.6 estabelece que uma transformação T num espaço V de dimensão finita é hermitica ou hemi-hermitica conforme a sua matriz relativa a uma base ortonormada é hermitica ou hemi-hermitica.

Estas matrizes podem introduzir-se doutro modo. Seja \bar{A} a matriz obtida pela substituição de cada elemento de A pelo seu complexo conjugado. A matriz \bar{A} diz-se a matriz *conjugada* de A . A matriz A é hermitica se e só se é igual à transposta da sua conjugada, $A = \bar{A}^t$. A matriz é hemi-hermitica se $A = -\bar{A}^t$.

A transposta da conjugada recebe uma designação especial.

DEFINIÇÃO DA MATRIZ ASSOCIADA (TRANSCONJUGADA) DE UMA MATRIZ. Dada uma matriz qualquer A , a transposta da conjugada, \bar{A}^t , designa-se por *associada* (ou *transconjugada*) de A e representa-se por A^* .

Portanto, uma matriz quadrada A é hermitica se $A = A^*$ e hemi-hermitica se $A = -A^*$.

Nota: A maior parte da antiga literatura relativa a matrizes utiliza a denominação de *adjunta* para a matriz complementos algébricos, que é um ente completamente distinto. A definição dada aqui concorda com a nomenclatura actual da teoria de operadores lineares.

5.9. Diagonalização de uma matriz hermitica ou hemi-hermitica.

TEOREMA 5.7. Toda a matriz A , $n \times n$, hermitica ou hemi-hermitica é semelhante à matriz diagonal $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ dos seus valores próprios. Além disso, tem-se

$$\Lambda = C^{-1}AC,$$

com C uma matriz não singular cuja inversa é a sua associada, $C^{-1} = C^*$.

Demonstração. Seja V o espaço dos n -tuplos de números complexos, e seja (e_1, \dots, e_n) a base ortonormada de vectores coordenados unitários. Se $x = \sum x_i e_i$ e $y = \sum y_i e_i$, consideremos o produto interno definido por $(x, y) = \sum x_i \bar{y}_i$. Para a matriz dada A , seja T a transformação representada por A relativa à base escolhida. Então

o teorema 5.4 diz-nos que V admite uma base ortonormada de vectores próprios (u_1, \dots, u_n) , a respeito da qual T tem uma representação em matriz diagonal $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, sendo λ_k o valor próprio pertencente a u_k . Posto que tanto A como Λ representam T elas são semelhantes, pelo que se terá $\Lambda = C^{-1}AC$, onde $C = (e_{ij})$ é a matriz não singular relacionando as duas bases:

$$[u_1, \dots, u_n] = [e_1, \dots, e_n]C.$$

Esta equação mostra que a coluna de ordem j de C é formada pelas componentes de u_j relativas a (e_1, \dots, e_n) . Deste modo c_{ij} é a componente de ordem i de u_j . O produto interno de u_j e u_i é dado por

$$(u_j, u_i) = \sum_{k=1}^n c_{kj} \bar{c}_{ki}.$$

Visto que $\{u_1, \dots, u_n\}$ é um conjunto ortonormado, isto mostra que $CC^* = I$, pelo que $C^{-1} = C^*$.

Nota: A demonstração do teorema 5.7 diz-nos também qual a maneira de determinar a matriz diagonalizadora C . Determinamos um conjunto ortonormado de vectores próprios u_1, \dots, u_n e depois utilizamos as componentes de u_j (relativas à base formada pelos vectores coordenados unitários) como elementos da coluna de ordem j de C .

EXEMPLO 1. A matriz real hermitica $A = \begin{bmatrix} 2 & 2 \\ 2 & 5 \end{bmatrix}$ tem os valores próprios $\lambda_1 = 1$ e $\lambda_2 = 6$. Os vectores próprios correspondentes a 1 são $t(2, -1)$, $t \neq 0$. Os correspondentes a 6 são $t(1, 2)$, $t \neq 0$. Os dois vectores próprios $u_1 = t(2, -1)$ e $u_2 = t(1, 2)$ com $t = 1/\sqrt{5}$ formam um conjunto ortonormado. Deste modo a matriz

$$C = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}$$

é a matriz diagonalizadora de A . Neste caso $C^* = C^T$ visto que t é real.

Verifica-se com facilidade que $C^*AC = \begin{bmatrix} 1 & 0 \\ 0 & 6 \end{bmatrix}$.

EXEMPLO 2. Se A é já uma matriz diagonal, então a matriz diagonalizadora C do teorema 5.7 ou deixa A invariável ou simplesmente reordena os elementos diagonais.

5.10. Matrizes unitárias. Matrizes ortogonais

DEFINIÇÃO. Uma matriz quadrada A diz-se unitária se $AA^* = I$. A matriz diz-se ortogonal se $AA^T = I$.

Nota: Toda a matriz real unitária é ortogonal visto que $A^* = A^T$.

O teorema 5.7 diz-nos que uma matriz hermitica ou hemi-hermitica pode sempre ser diagonalizada por meio de uma matriz unitária. Uma matriz real, hermitica, tem valores próprios reais e os vectores próprios correspondentes podem tornar-se reais. Portanto uma matriz hermitica real pode ser diagonalizada por uma matriz real ortogonal. Tal não é *verdadeiro* para as matrizes hemi-hermiticas reais (Ver Exercício 11 da Secção 5.11).

Assim temos os conceitos seguintes

DEFINIÇÃO. Uma matriz quadrada A com elementos reais ou complexos diz-se *simétrica* se $A = A^t$; diz-se *hemi-simétrica* se $A = -A^t$.

EXEMPLO 3. Se A é real, a sua associada é igual à sua transposta $A^* = A^t$. Assim, toda a matriz hermitica *real* é simétrica, mas uma matriz simétrica não será necessariamente hermitica.

EXEMPLO 4. Se $A = \begin{bmatrix} 1+i & 2 \\ 3-i & 4i \end{bmatrix}$, então $\bar{A} = \begin{bmatrix} 1-i & 2 \\ 3+i & -4i \end{bmatrix}$, $A^t = \begin{bmatrix} 1+i & 3-i \\ 2 & 4i \end{bmatrix}$
e $A^* = \begin{bmatrix} 1-i & 3+i \\ 2 & -4i \end{bmatrix}$.

EXEMPLO 5. Ambas as matrizes $\begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$ e $\begin{bmatrix} 1 & 2+i \\ 2-i & 3 \end{bmatrix}$ são hermiticas. A primeira é simétrica, e a segunda não.

EXEMPLO 6. Ambas as matrizes $\begin{bmatrix} 0 & -2 \\ 2 & 0 \end{bmatrix}$ e $\begin{bmatrix} i & -2 \\ 2 & 3i \end{bmatrix}$ são hemi-hermiticas. A primeira é hemi-simétrica, e a segunda não.

EXEMPLO 7. Todos os elementos diagonais de uma matriz hermitica são reais. Todos os elementos diagonais de uma matriz hemi-hermitica são imaginários puros. Todos os elementos diagonais de uma matriz hemi-simétrica são zeros.

EXEMPLO 8. Para uma matriz quadrada qualquer A , a matriz $B = \frac{1}{2}(A + A^*)$ é hermitica, e a matriz $C = \frac{1}{2}(A - A^*)$ é hemi-hermitica. A sua soma é A . Assim, toda a matriz quadrada A pode ser expressa por uma soma $A = B + C$, com B uma matriz hermitica e C uma matriz hemi-hermitica. É um exercício simples a verificação de que a decomposição é única. Também toda a matriz quadrada A pode ser expressa, de uma maneira única, como a soma de uma matriz simétrica, $\frac{1}{2}(A + A^t)$, com uma matriz hemi-simétrica, $\frac{1}{2}(A - A^t)$.

EXEMPLO 9. Se A é ortogonal tem-se $1 = \det(AA^t) = (\det A)(\det A^t) = (\det A)^2$, pelo que $\det A = \pm 1$.

5.11. Exercícios

1. Determinar quais das seguintes matrizes são simétricas, hemi-simétricas, hermiticas, hemi-hermiticas.

$$(a) \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 2 & 3 & 4 \end{bmatrix}, \quad (b) \begin{bmatrix} 0 & i & 2 \\ i & 0 & 3 \\ -2 & -3 & 4i \end{bmatrix}, \quad (c) \begin{bmatrix} 0 & i & 2 \\ -i & 0 & 3 \\ -2 & -3 & 0 \end{bmatrix}, \quad (d) \begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & 3 \\ -2 & -3 & 0 \end{bmatrix}.$$

2. (a) Verificar que a matriz 2×2 , $A = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$ é uma matriz ortogonal.

(b) Seja T a transformação linear com a matriz anterior A relativa à base ordinária $\{i, j\}$. Provar que T aplica cada ponto do plano com coordenadas polares (r, α) sobre o ponto com coordenadas polares $(r, \alpha + \theta)$. Assim, T é uma rotação do plano em torno da origem, sendo θ ângulo de rotação.

3. Seja V o espaço real tridimensional com a base usual i, j, k . Provar que cada uma das seguintes matrizes é ortogonal e representa a transformação indicada.

$$(a) \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad (\text{simetria no plano } XOY)$$

$$(b) \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad (\text{simetria em relação ao eixo } OX)$$

$$(c) \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad (\text{simetria a respeito da origem})$$

$$(d) \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{bmatrix} \quad (\text{rotação em torno de } OX)$$

$$(e) \begin{bmatrix} -1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{bmatrix} \quad (\text{rotação em torno de } OX \text{ seguida de simetria em relação ao plano } YOZ).$$

4. Uma matriz ortogonal real A diz-se *própria* se $\det A = 1$ e *imprópria* se $\det A = -1$.

(a) Se A é uma matriz 2×2 própria, provar que $A = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$ para algum θ . Esta representa uma rotação de um ângulo θ .

(b) Provar que $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ e $\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$ são matrizes impróprias. A primeira representa uma simetria no plano XOY em relação a OX ; a segunda representa uma simetria em relação ao eixo OY . Determinar todas as matrizes 2×2 impróprias.

Em cada um dos Exercícios 5 a 8 determinar: (a) um conjunto ortogonal dos vectores próprios para A , e (b) uma matriz unitária C tal que $C^{-1}AC$ seja uma matriz diagonal.

$$5. A = \begin{bmatrix} 9 & 12 \\ 12 & 16 \end{bmatrix}.$$

$$6. A = \begin{bmatrix} 0 & -2 \\ 2 & 0 \end{bmatrix}.$$

$$7. A = \begin{bmatrix} 1 & 3 & 0 \\ 3 & -2 & -1 \\ 0 & -1 & 1 \end{bmatrix}.$$

$$8. A = \begin{bmatrix} 1 & 3 & 4 \\ 3 & 1 & 0 \\ 4 & 0 & 1 \end{bmatrix}.$$

9. Dizer quais das matrizes seguintes são unitárias e quais são ortogonais (a, b, θ são reais).

$$(a) \begin{bmatrix} e^{ia} & 0 \\ 0 & e^{ib} \end{bmatrix}, \quad (b) \begin{bmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{bmatrix}, \quad (c) \begin{bmatrix} \frac{1}{2}\sqrt{2} & -\frac{1}{3}\sqrt{3} & \frac{1}{6}\sqrt{6} \\ 0 & \frac{1}{3}\sqrt{3} & \frac{1}{3}\sqrt{6} \\ \frac{1}{2}\sqrt{2} & \frac{1}{3}\sqrt{3} & -\frac{1}{6}\sqrt{6} \end{bmatrix}.$$

10. A teoria da relatividade restrita utiliza as equações

$$x' = a(x - vt), \quad y' = y, \quad z' = z, \quad t' = a(t - vx/c^2).$$

Aqui v representa a velocidade dum objecto em movimento, c a velocidade da luz, e $a = c/\sqrt{c^2 - v^2}$. A transformação linear que aplica (x, y, z, t) em (x', y', z', t') diz-se uma *transformação de Lorentz*.

(a) Sejam $(x_1, x_2, x_3, x_4) = (x, y, z, ict)$ e $(x'_1, x'_2, x'_3, x'_4) = (x', y', z', ict')$. Mostrar que as quatro equações podem escrever-se na forma de uma equação matricial

$$[x'_1, x'_2, x'_3, x'_4] = [x_1, x_2, x_3, x_4] \begin{bmatrix} a & 0 & 0 & -iav/c \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ iav/c & 0 & 0 & a \end{bmatrix}.$$

(b) Provar que a matriz 4×4 da alínea (a) é ortogonal mas não unitária.

11. Seja a um número real diferente de zero e A a matriz hemi-simétrica $A = \begin{bmatrix} 0 & a \\ -a & 0 \end{bmatrix}$.

(a) Determinar um conjunto ortonormado de vectores próprios de A .

(b) Determinar uma matriz unitária C tal que $C^{-1}AC$ é uma matriz diagonal.

(c) Provar que não existe nenhuma matriz ortogonal C tal que $C^{-1}AC$ seja uma matriz diagonal.

12. Se os valores próprios de uma matriz A hermitica ou hemi-hermitica são todos iguais a c , provar que $A = cI$.

13. Se A é uma matriz real hemi-simétrica, provar que ambas as matrizes $I - A$ e $I + A$ são não singulares e que $(I - A)(I + A)^{-1}$ é ortogonal.

14. Para cada uma das proposições seguintes relativas a matrizes $n \times n$, efectuar a correspondente demonstração ou apresentar um contra exemplo.

(a) Se A e B são unitárias, então $A + B$ é unitária.

(b) Se A e B são unitárias, então AB é unitária.

(c) Se A e AB são unitárias, então B é unitária.

(d) Se A e B são unitárias, então $A + B$ não é unitária.

5.12. Formas quadráticas

Seja V o espaço euclidiano real e seja $T: V \rightarrow V$ um operador simétrico. Isto significa que T pode ser mudado de um para outro factor no produto interno

$$(T(x), y) = (x, T(y)) \quad \text{para quaisquer } x \text{ e } y \text{ em } V$$

Dado T , definimos uma função real Q em V pela equação

$$Q(x) = (T(x), x).$$

A função Q chama-se a *forma quadrática* associada com T . A designação “quadrática” é sugerida pelo teorema seguinte que mostra que, no caso de dimensão finita, $Q(x)$ é um polinómio do segundo grau nas componentes de x .

TEOREMA 5.8. *Sejam (e_1, \dots, e_n) uma base ortonormada para um espaço euclidiano real V , $T: V \rightarrow V$ uma transformação simétrica e $A = (a_{ij})$ a matriz de T relativa àquela base. A forma quadrática $(Q(x) = (T(x), x))$ está relacionada com A do modo seguinte:*

$$(5.7) \quad Q(x) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j \quad \text{se } x = \sum_{i=1}^n x_i e_i.$$

Demonstração. Devido à linearidade temos $T(x) = \sum x_i T(e_i)$. Portanto

$$Q(x) = \left(\sum_{i=1}^n x_i T(e_i), \sum_{j=1}^n x_j e_j \right) = \sum_{i=1}^n \sum_{j=1}^n x_i x_j (T(e_i), e_j).$$

igualdade que demonstra (5.7) já que $a_{ij} = \bar{a}_{ji} = (T(e_i), e_j)$.

A soma que aparece em (5.7) é ainda provida de significado mesmo se a matriz A é não simétrica.

DEFINIÇÃO. *Seja V um espaço euclidiano real qualquer, com uma base ortonormada (e_1, \dots, e_n) e $A = (a_{ij})$ qualquer matriz $n \times n$ de escalares. A função de valores escalares Q definida em cada elemento $x = \sum x_i e_i$ de V pela dupla soma*

$$(5.8) \quad Q(x) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j$$

diz-se uma forma quadrática associada com A .

Se A é uma matriz diagonal, então $a_{ij} = 0$ se $i \neq j$ pelo que a soma em (5.8) contém unicamente termos quadráticos e pode escrever-se mais simplesmente na forma

$$Q(x) = \sum_{i=1}^n a_{ii} x_i^2.$$

Neste caso a forma quadrática diz-se uma *forma diagonal*.

A dupla soma em (5.8) também pode exprimir-se como um produto de três matrizes.

TEOREMA 5.9. Se $X = [x_1, \dots, x_n]$ é uma matriz linha $1 \times n$ e $A = (a_{ij})$ é uma matriz $n \times n$, então XAX^t é uma matriz 1×1 cujo elemento é

$$(5.9) \quad \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j.$$

Demonstração. O produto XA é uma matriz $1 \times n$, $XA = [y_1, \dots, y_n]$ cujo elemento y_j é o produto escalar de X pela coluna j de A ,

$$y_j = \sum_{i=1}^n x_i a_{ij}.$$

Portanto o produto XAX^t é uma matriz 1×1 cujo único elemento é o produto escalar

$$\sum_{j=1}^n y_j x_j = \sum_{j=1}^n \left(\sum_{i=1}^n x_i a_{ij} \right) x_j = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j.$$

Nota: É usual identificar a matriz 1×1 , XAX^t , com a soma (5.9) e chamar o produto XAX^t uma forma quadrática. A equação (5.8) escreve-se mais simplesmente na forma seguinte:

$$Q(x) = XAX^t.$$

EXEMPLO 1. $A = \begin{bmatrix} 1 & -1 \\ -3 & 5 \end{bmatrix}$, $X = [x_1, x_2]$. Então temos

$$XA = [x_1, x_2] \begin{bmatrix} 1 & -1 \\ -3 & 5 \end{bmatrix} = [x_1 - 3x_2, -x_1 + 5x_2],$$

e por conseguinte

$$XAX^t = [x_1 - 3x_2, -x_1 + 5x_2] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1^2 - 3x_2x_1 - x_1x_2 + 5x_2^2.$$

EXEMPLO 2. Seja $B = \begin{bmatrix} 1 & -2 \\ -2 & 5 \end{bmatrix}$, $X = [x_1, x_2]$. Então temos

$$XBX^t = [x_1, x_2] \begin{bmatrix} 1 & -2 \\ -2 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1^2 - 2x_2x_1 - 2x_1x_2 + 5x_2^2.$$

Em ambos os Exemplos 1 e 2 os termos rectangulares dão, ao somarem-se, $-4x_1x_2$ pelo que $XAX^t = XBX^t$. Concluimos assim que matrizes diferentes podem conduzir à mesma forma quadrática. Observe-se que, pelo menos uma dessas matrizes é simétrica. O que é um exemplo de seguinte teorema:

TEOREMA 5.10. Para qualquer matriz A , $n \times n$, e qualquer matriz linha X , $1 \times n$, tem-se $XAX^t = XBX^t$ com B a matriz simétrica $B = \frac{1}{2}(A + A^t)$.

Demonstração. Porque XAX^t é uma matriz 1×1 ela é igual à sua transposta, $XAX^t = (XAX^t)^t$. Mas a transposta de um produto de matrizes é o produto das transpostas das matrizes factores consideradas por ordem inversa, pelo que se tem $(XAX^t)^t = XA^tX^t$. Deste modo $XAX^t = \frac{1}{2}XAX^t + \frac{1}{2}XA^tX^t = XBX^t$.

5.13. Redução de uma forma quadrática real à forma diagonal

Uma matriz simétrica real A é hermitica. Portanto, pelo teorema 5.7, é semelhante à matriz diagonal, $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, dos seus valores próprios. Além disso, temos $\Lambda = C^tAC$, onde C é uma matriz ortogonal. Vamos agora demonstrar que C pode utilizar-se para converter a forma quadrática XAX^t numa forma diagonal.

TEOREMA 5.11. Se XAX^t é uma forma quadrática associada com uma matriz simétrica real A e se C é uma matriz ortogonal que converte A numa matriz diagonal $\Lambda = C^tAC$, então tem-se

$$XAX^t = Y\Lambda Y^t = \sum_{i=1}^n \lambda_i y_i^2,$$

onde $Y = [y_1, \dots, y_n]$ é a matriz linha $Y = XC$, e $\lambda_1, \dots, \lambda_n$ são os valores próprios de A .

Demonstração. Uma vez que C é ortogonal temos $C^{-1} = C^t$. Deste modo a equação $Y = XC$ implica $X = YC^t$, e obtemos

$$XAX^t = (YC^t)A(YC^t)^t = Y(C^tAC)Y^t = Y\Lambda Y^t.$$

Nota: O teorema 5.11 exprime que a transformação linear $Y = XC$ reduz a forma quadrática XAX^t à forma diagonal $Y\Lambda Y^t$.

EXEMPLO 1. A forma quadrática correspondente à matriz identidade é

$$XIX^t = \sum_{i=1}^n x_i^2 = \|X\|^2,$$

o quadrado do comprimento do vector $X = (x_1, \dots, x_n)$. Uma transformação linear $Y = XC$, onde C é uma matriz ortogonal, define uma nova forma quadrática $Y\Lambda Y^t$ com $\Lambda = C^tAC = CC^t = I$. Uma vez que $XIX^t = YIY^t$ temos $\|X\|^2 = \|Y\|^2$, donde resulta que Y tem o mesmo comprimento que X . Uma transformação linear que preserva o comprimento de cada vector diz-se uma *isometria*. Estas transformações serão analisadas mais em pormenor na Seção 5.19.

EXEMPLO 2. Determinar uma matriz ortogonal C que reduza a forma quadrática $Q(x) = 2x_1^2 + 4x_1x_2 + 5x_2^2$ à forma diagonal.

Resolução. Escrevemos $Q(x) = XAX'$, com $A = \begin{bmatrix} 2 & 2 \\ 2 & 5 \end{bmatrix}$. Esta matriz simétrica foi diagonalizada no Exemplo 1 a seguir ao teorema 5.7. Ela tem os valores próprios $\lambda_1 = 1$, $\lambda_2 = 6$, e um conjunto ortonormado de vectores próprios u_1 e u_2 , $u_1 = t(2, -1)$, $u_2 = t(1, 2)$, $t = 1/\sqrt{5}$. Uma matriz diagonalizadora é $C = t \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}$. A correspondente forma diagonal é

$$YAY' = \lambda_1 y_1^2 + \lambda_2 y_2^2 = y_1^2 + 6y_2^2.$$

O resultado deste exemplo é susceptível de uma interpretação geométrica simples, representado na figura 5.1. A transformação linear $Y = XC$ pode considerar-se como uma rotação que aplica os vectores da base, i e j , sobre os novos vectores u_1 , u_2 da base. Um ponto de coordenadas (x_1, x_2) em relação à primeira base admita novas coordenadas (y_1, y_2) em relação à segunda. Visto que $XAX' = YAY'$, o conjunto de pontos (x_1, x_2) satisfazendo à equação $XAX' = c$, para algum c , é idêntico ao conjunto de pontos (y_1, y_2) satisfazendo a $YAY' = c$. A segunda equação escrita na forma $y_1^2 + 6y_2^2 = c$ é a equação cartesiana de uma elipse se $c > 0$. Portanto a equação $XAX' = c$, posta na forma, $2x_1^2 + 4x_1x_2 + 5x_2^2 = c$, representa a mesma elipse no sistema de coordenadas inicial. A figura 5.1 mostra a elipse correspondente a $c = 9$.

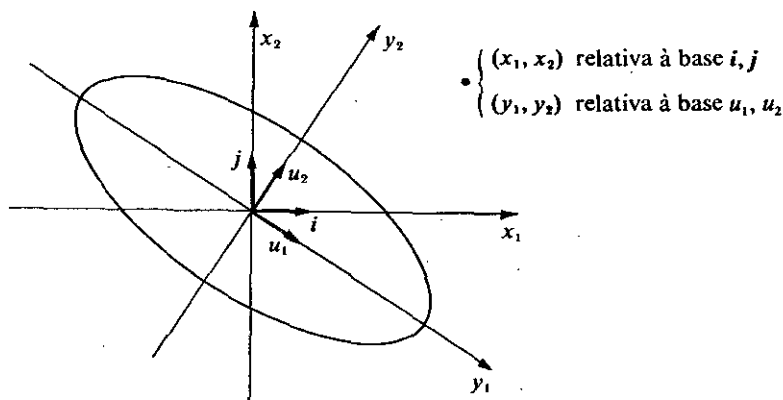


FIG. 5.1. Rotação dos eixos por uma matriz ortogonal. A elipse admite a equação cartesiana $XAX' = 9$ no sistema X_1OX_2 , e a equação $YAY' = 9$ no sistema Y_1OY_2 .

5.14. Aplicações à geometria analítica

A redução de uma forma quadrática à forma diagonal pode ser utilizada para identificar o conjunto de todos os pontos (x, y) no plano que satisfazem a uma equação cartesiana da forma

$$(5.10) \quad ax^2 + bxy + cy^2 + dx + ey + f = 0.$$

Verificaremos que este conjunto é sempre uma cónica, isto é, uma elipse, uma hipérbole ou uma parábola ou um dos casos degenerados (o conjunto vazio, um único ponto, ou uma ou duas rectas). O tipo de cónica é definido pelos termos do segundo grau, isto é, pela forma quadrática $ax^2 + bxy + cy^2$. Para estarmos de acordo com a notação usada antes, escrevemos x_1 para x , x_2 em vez de y , e exprimimos esta forma quadrática como um produto matricial,

$$XAX^t = ax_1^2 + bx_1x_2 + cx_2^2,$$

onde $X = [x_1, x_2]$ e $A = \begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix}$. Por uma rotação $Y = XC$ reduzimos esta forma quadrática à forma diagonal $\lambda_1 y_1^2 + \lambda_2 y_2^2$, onde λ_1 e λ_2 são os valores próprios de A . Um conjunto ortonormado de vectores próprios u_1 e u_2 determina um novo conjunto de eixos coordenados, relativamente aos quais a equação cartesiana (5.10) vem

$$(5.11) \quad \lambda_1 y_1^2 + \lambda_2 y_2^2 + d'y_1 + e'y_2 + f = 0,$$

com novos coeficientes d' e e' nos termos lineares. Nesta equação não há termos rectangulares da forma $y_1 y_2$, pelo que o tipo de cónica se identificará pela análise dos vectores próprios λ_1 e λ_2 . Se a cónica é não degenerada, a equação (5.11) representa uma *elipse* se λ_1, λ_2 têm o mesmo sinal, uma *hipérbole* se λ_1 e λ_2 têm sinais contrários e uma *parábola* se qualquer dos valores λ_1 e λ_2 for nulo. Os três casos correspondem a $\lambda_1 \lambda_2 > 0$, $\lambda_1 \lambda_2 < 0$, e $\lambda_1 \lambda_2 = 0$. Vamos apresentar alguns exemplos específicos.

EXEMPLO 1. $2x^2 + 4xy + 5y^2 + 4x + 13y - \frac{1}{4} = 0$. Escrevemos novamente

$$(5.12) \quad 2x_1^2 + 4x_1x_2 + 5x_2^2 + 4x_1 + 13x_2 - \frac{1}{4} = 0.$$

A forma quadrática $2x_1^2 + 4x_1x_2 + 5x_2^2$ é uma das analisadas no Exemplo 2 da Secção anterior. A sua matriz tem valores próprios $\lambda_1 = 1$, $\lambda_2 = 6$ e um conjunto ortonormado $u_1 = t(2, -1)$, $u_2 = t(1, 2)$ com $t = 1/\sqrt{5}$. Uma matriz ortogonal diagonalizante é $C = t \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}$. Esta reduz a parte quadrática de (5.12) à forma $y_1^2 + 6y_2^2$. Para determinar o efeito sobre a parte linear escrevemos a equação da rotação $Y = XC$ na forma $X = YC^t$ e obtemos

$$[x_1, x_2] = \frac{1}{\sqrt{5}} [y_1, y_2] \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}, \quad x_1 = \frac{1}{\sqrt{5}} (2y_1 + y_2), \quad x_2 = \frac{1}{\sqrt{5}} (-y_1 + 2y_2).$$

Portanto a parte linear $4x_1 + 13x_2$ transforma-se em

$$\frac{4}{\sqrt{5}} (2y_1 + y_2) + \frac{13}{\sqrt{5}} (-y_1 + 2y_2) = -\sqrt{5} y_1 + 6\sqrt{5} y_2.$$

A equação cartesiana transformada vem

$$y_1^2 + 6y_2^2 - \sqrt{5} y_1 + 6\sqrt{5} y_2 - \frac{1}{4} = 0.$$

Completando os quadros em y_1 e y_2 escrevemos de novo a equação na forma seguinte

$$(y_1 - \frac{1}{2}\sqrt{5})^2 + 6(y_2 + \frac{1}{2}\sqrt{5})^2 = 9.$$

a qual é a equação de uma elipse cujo centro é o ponto $(\frac{1}{2}\sqrt{5}, -\frac{1}{2}\sqrt{5})$ no sistema coordenado OY_1OY_2 . Os sentidos positivos de Oy_1 e Oy_2 são determinados pelos vectores próprios u_1 e u_2 , como se indica na figura 5.2.

Podemos simplificar a equação um pouco mais escrevendo

$$z_1 = y_1 - \frac{1}{2}\sqrt{5}, \quad z_2 = y_2 + \frac{1}{2}\sqrt{5}.$$

Geometricamente isto é equivalente a definir um novo sistema de eixos coordenados paralelos aos eixos do sistema OY_1OY_2 , mas com a origem no centro da elipse. No sistema $O'Z_1Z_2$ a equação da elipse escreve-se simplesmente

$$z_1^2 + 6z_2^2 = 9, \quad \text{ou} \quad \frac{z_1^2}{9} + \frac{z_2^2}{3/2} = 1.$$

Na figura 5.2 está traçada a elipse e os três sistemas de eixos coordenados.

EXEMPLO 2. $2x^2 - 4xy - y^2 - 4x + 10y - 13 = 0$. Escrevamos a equação na forma

$$2x_1^2 - 4x_1x_2 - x_2^2 - 4x_1 + 10x_2 - 13 = 0.$$

A parte quadrática e $XA X^t$, com $A = \begin{bmatrix} 2 & -2 \\ -2 & -1 \end{bmatrix}$. Esta matriz tem os valores próprios

$\lambda_1 = 3$, $\lambda_2 = -2$. Um conjunto ortonormado de vectores próprios é $u_1 = t(2, -1)$,

$u_2 = t(1, 2)$, com $t = 1/\sqrt{5}$. Uma matriz diagonalizadora ortogonal é $C = t \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}$.

A equação da rotação $X = YC$ vem

$$x_1 = \frac{1}{\sqrt{5}} (2y_1 + y_2), \quad x_2 = \frac{1}{\sqrt{5}} (-y_1 + 2y_2).$$

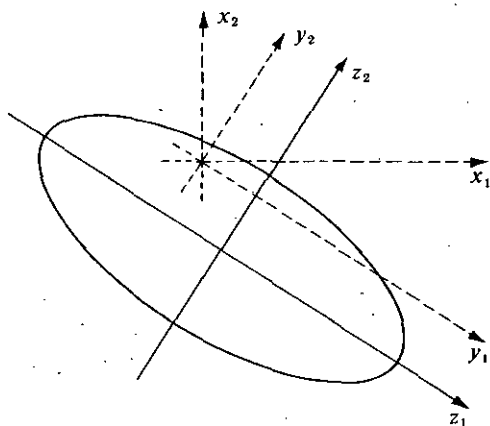


FIG. 5.2. Rotação e translação de eixos coordenados. A rotação $Y = XC$ é seguida da translação $z_1 = y_1 - \frac{1}{3}\sqrt{5}$, $z_2 = y_2 + \frac{4}{3}\sqrt{5}$.

Portanto as equações transformadas vem

$$3y_1^2 - 2y_2^2 - \frac{4}{\sqrt{5}}(2y_1 + y_2) + \frac{10}{\sqrt{5}}(-y_1 + 2y_2) - 13 = 0,$$

ou

$$3y_1^2 - 2y_2^2 - \frac{18}{\sqrt{5}}y_1 + \frac{16}{\sqrt{5}}y_2 - 13 = 0.$$

Completando os quadrados em y_1 e y_2 obtemos a equação

$$3(y_1 - \frac{3}{\sqrt{5}})^2 - 2(y_2 - \frac{4}{\sqrt{5}})^2 = 12,$$

a qual representa uma hipérbole com centro em $(\frac{3}{\sqrt{5}}, \frac{4}{\sqrt{5}})$ no sistema OY_1Y_2 . A translação $z_1 = y_1 - \frac{3}{\sqrt{5}}$, $z_2 = y_2 - \frac{4}{\sqrt{5}}$ simplifica esta equação para

$$3z_1^2 - 2z_2^2 = 12, \quad \text{ou} \quad \frac{z_1^2}{4} - \frac{z_2^2}{6} = 1.$$

A hipérbole está traçada na fig. 5.3 (a). Os vectores próprios u_1 e u_2 determinam os semi-eixos positivos OY_1 e OY_2 .

EXEMPLO 3. $9x^2 + 24xy + 16y^2 - 20x + 15y = 0$. Escrevemos de novo a equação na forma

$$9x_1^2 + 24x_1x_2 + 16x_2^2 - 20x_1 + 15x_2 = 0.$$

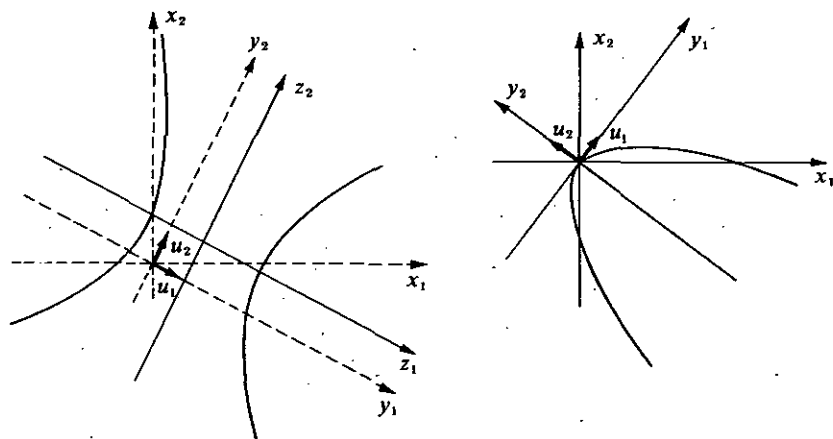
(a) Hipérbole: $3z_1^2 - 2z_2^2 = 12$ (b) Parábola: $y_1^2 + y_2 = 0$

FIG. 5.3. As curvas dos Exemplos 2 e 3.

A matriz simétrica para a parte quadrática é $A = \begin{bmatrix} 9 & 12 \\ 12 & 16 \end{bmatrix}$. Os seus valores próprios são $\lambda_1 = 25$, $\lambda_2 = 0$. Um conjunto ortonormado de vectores próprios é $u_1 = \frac{1}{5}(3, 4)$, $u_2 = \frac{1}{5}(-4, 3)$. Uma matriz ortogonal diagonalizadora é $C = \frac{1}{5} \begin{bmatrix} 3 & -4 \\ 4 & 3 \end{bmatrix}$. A equação da rotação $X = YC'$ dá-nos

$$x_1 = \frac{1}{5}(3y_1 - 4y_2), \quad x_2 = \frac{1}{5}(4y_1 + 3y_2).$$

Portanto a equação cartesiana transformada escreve-se

$$25y_1^2 - \frac{29}{5}(3y_1 - 4y_2) + \frac{15}{5}(4y_1 + 3y_2) = 0.$$

Simplificando esta equação chega-se a $y_1^2 + y_2 = 0$, a equação de uma parábola com vértice na origem. A parábola está traçada na fig. 5.3 (b).

EXEMPLO 4. Casos degenerados. O conhecimento unicamente dos valores próprios não revela se a equação cartesiana representa uma cónica degenerada. Por exemplo, as três equações $x^2 + 2y^2 = 1$, $x^2 + 2y^2 = 0$, e $x^2 + 2y^2 = -1$ admitem todas os mesmos valores próprios; a primeira representa uma elipse não degenerada, a segunda só é verificada para $(x, y) = (0, 0)$ e a terceira representa o conjunto vazio. As duas últimas podem considerar-se como casos degenerados da elipse.

O gráfico da equação $y^2 = 0$ é o eixo OX . A equação $y^2 - 1 = 0$ representa duas rectas paralelas $y = 1$ e $y = -1$. Estas podem considerarse como casos degenerados da parábola. A equação $x^2 - 4y^2 = 0$ representa duas rectas que se intersectam já que é satis-

feita se $x - 2y = 0$ ou se $x + 2y = 0$, e pode portanto considerar-se como um caso degenerado da hipérbole.

Porém, se a equação cartesiana $ax^2 + bxy + cy^2 + dx + ey + f = 0$ representa uma cónica não degenerada, então o tipo de cónica pode ser determinado de modo muito fácil. O polinómio característico da matriz da forma quadrática $ax^2 + bxy + cy^2$ é

$$\det \begin{bmatrix} \lambda - a & -b/2 \\ -b/2 & \lambda - c \end{bmatrix} = \lambda^2 - (a + c)\lambda + (ac - \frac{1}{4}b^2) = (\lambda - \lambda_1)(\lambda - \lambda_2).$$

Portanto o produto dos seus valores próprios é

$$\lambda_1 \lambda_2 = ac - \frac{1}{4}b^2 = \frac{1}{4}(4ac - b^2).$$

Visto que o tipo de cónica se determina pelo sinal do produto $\lambda_1 \lambda_2$, vemos que a cónica é uma *elipse*, *hipérbole* ou *parábola* conforme $4ac - b^2$ é *positivo*, *negativo* ou *nulo*. O número $4ac - b^2$ diz-se o *discriminante* da forma quadrática $ax^2 + bxy + cy^2$. Nos Exemplos 1, 2 e 3 o discriminante tem os valores 34, -24 e 0, respectivamente.

5.15. Exercícios

Em cada um dos Exercícios 1 a 7, determinar: (a) uma matriz simétrica A para a forma quadrática; (b) os valores próprios de A ; (c) um conjunto ortonormado de vectores próprios; (d) uma matriz diagonalizadora ortogonal C .

1. $4x_1^2 + 4x_1x_2 + x_2^2$.

2. x_1x_2 .

3. $x_1^2 + 2x_1x_2 - x_2^2$.

4. $34x_1^2 - 24x_1x_2 + 41x_2^2$.

5. $x_1^2 + x_1x_2 + x_1x_3 + x_2x_3$.

6. $2x_1^2 + 4x_1x_3 + x_2^2 - x_3^2$.

7. $3x_1^2 + 4x_1x_2 + 8x_1x_3 + 4x_2x_3 + 3x_3^2$.

Em cada um dos Exercícios 8 a 18, identificar a cónica definida pela respectiva equação e desenhá-la.

8. $y^2 - 2xy + 2x^2 - 5 = 0$.

9. $y^2 - 2xy + 5x = 0$.

10. $y^2 - 2xy + x^2 - 5x = 0$.

11. $5x^2 - 4xy + 2y^2 - 6 = 0$.

12. $19x^2 + 4xy + 16y^2 - 212x + 104y = 356$.

13. $9x^2 + 24xy + 16y^2 - 52x + 14y = 6$.

14. $5x^2 + 6xy + 5y^2 - 2 = 0$.

15. $x^2 + 2xy + y^2 - 2x + 2y + 3 = 0$.

16. $2x^2 + 4xy + 5y^2 - 2x - y - 4 = 0$.

17. $x^2 + 4xy - 2y^2 - 12 = 0$.

18. $xy + y - 2x - 2 = 0$.

19. Para que valor (ou valores) de c será o gráfico da equação $2xy - 4x + 7y + c = 0$ um par de rectas?

20. Se a equação $ax^2 + bxy + cy^2 = 1$ representa uma elipse, provar que a área da região que ela limita é $2\pi/\sqrt{4ac - b^2}$. Isto confere ao discriminante $4ac - b^2$ um significado geométrico.

★ 5.16† Valores próprios de uma transformação simétrica obtidos como valores de sua forma quadrática

Prescindindo agora da exigência de que V seja de dimensão finita, vamos encontrar uma relação entre os valores próprios de um operador simétrico e a sua forma quadrática.

Suponhamos que x é um vector próprio com norma 1 correspondendo a um valor próprio λ . Então $T(x) = \lambda x$ pelo que se tem

$$(5.13) \quad Q(x) = (T(x), x) = (\lambda x, x) = \lambda(x, x) = \lambda,$$

visto que $(x, x) = 1$. O conjunto de todos os x de V satisfazendo a $(x, x) = 1$ chama-se a *esfera unitária* de V . A equação (5.13) demonstra o seguinte teorema

TEOREMA 5.12. *Se $T: V \rightarrow V$ é uma transformação simétrica num espaço euclidiano real V e $Q(x) = (T(x), x)$, então os valores próprios de T (se existirem) encontram-se entre os valores que Q toma na esfera unitária de V .*

EXEMPLO. Seja $V = V_2(\mathbb{R})$ com a base usual (i, j) e o produto escalar como produto interno. Seja T uma transformação simétrica com matriz $A = \begin{bmatrix} 4 & 0 \\ 0 & 8 \end{bmatrix}$. Então a forma quadrática de T é dada por

$$Q(x) = \sum_{i=1}^2 \sum_{j=1}^2 a_{ij} x_i x_j = 4x_1^2 + 8x_2^2.$$

Os valores próprios de T são $\lambda_1 = 4$, $\lambda_2 = 8$. É fácil de ver que estes valores próprios são, respectivamente, os valores máximo e mínimo que Q toma no círculo unidade $x_1^2 + x_2^2 = 1$. Com efeito, neste círculo temos

$$Q(x) = 4(x_1^2 + x_2^2) + 4x_2^2 = 4 + 4x_2^2, \quad \text{onde } -1 \leq x_2 \leq 1.$$

Este alcança o seu valor mínimo, 4, quando $x_2 = 0$ e o seu valor máximo, 8, quando $x_2 = \pm 1$.

A figura 5.4 mostra o círculo unidade e duas elipses. A elipse interior tem a equação cartesiana $4x_1^2 + 8x_2^2 = 4$. É formada por todos os pontos $x = (x_1, x_2)$ no plano, satisfazendo $Q(x) = 4$. A elipse exterior tem por equação cartesiana $4x_1^2 + 8x_2^2 = 8$ e é formada por todos os pontos satisfazendo a $Q(x) = 8$. Os pontos $(\pm 1, 0)$, nos quais a elipse interior é tangente ao círculo unidade, são vectores próprios correspondentes ao valor

† As Secções assinaladas com asterisco podem suprimir-se, ou deixar-se para posterior leitura, sem perda de continuidade.

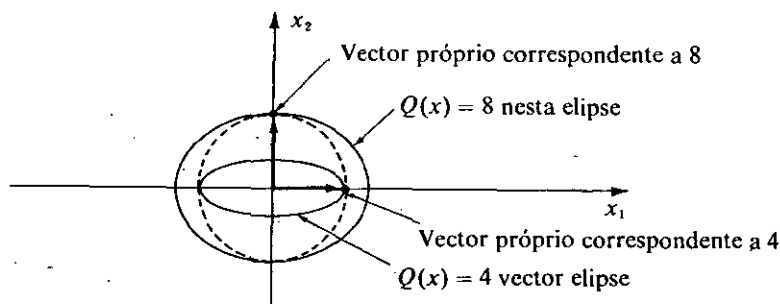


FIG. 5.4. Relação geométrica entre os valores próprios de T e os valores de Q na esfera unidade, ilustrada com um exemplo bidimensional.

próprio 4. Os pontos $(0, \pm 1)$ da elipse exterior são vectores próprios correspondentes ao valor próprio 8.

O exemplo anterior ilustra propriedades extremais dos valores próprios que são válidas com maior generalidade. Na secção que se segue provaremos que o menor e maior valor próprio (se existirem) são sempre os valores mínimo e máximo que Q toma na esfera unitária. Na discussão destas propriedades extremas faremos uso do seguinte teorema relativo a formas quadráticas, o qual não exige que V seja de dimensão finita.

TEOREMA 5.13. *Seja $T: V \rightarrow V$ uma transformação simétrica num espaço euclidiano real V dotado com a forma quadrática $Q(x) = (T(x), x)$. Admita-se que Q não muda de sinal em V . Então se $Q(x) = 0$ para algum x em V tem-se igualmente $T(x) = 0$. Por outras palavras, se Q não muda de sinal então Q é nulo unicamente no espaço nulo de T .*

Demonstração. Suponhamos $Q(x) = 0$ para algum x em V e seja y qualquer elemento de V . Escolhamos um escalar real t e consideremos $Q(x + ty)$. Aplicando a linearidade de T , linearidade do produto interno, e simetria de T , temos

$$\begin{aligned} Q(x + ty) &= (T(x + ty), x + ty) = (T(x) + tT(y), x + ty) \\ &= (T(x), x) + t(T(x), y) + t(T(y), x) + t^2(T(y), y) \\ &= Q(x) + 2t(T(x), y) + t^2Q(y) = at + bt^2, \end{aligned}$$

onde $a = 2(T(x), y)$ e $b = Q(y)$. Se Q é não negativa em V temos a desigualdade

$$at + bt^2 \geq 0 \text{ para todo real } t.$$

Por outras palavras, o polinómio quadrático $p(t) = at + bt^2$ tem o seu valor mínimo para $t = 0$. Consequentemente $p'(0) = 0$. Mas $p'(0) = a = 2(T(x), y)$, pelo que $(T(x),$

$y) = 0$. Uma vez que y era arbitrário, podemos em particular tomar $y = T(x)$, obtendo $(T(x), T(x)) = 0$. Isto prova que $T(x) = 0$.

Se Q é não positiva em V obtemos $p(t) = at + bt^2 \leq 0$ para todo t , pelo que p tem um máximo quando $t = 0$, e consequentemente $p'(0) = 0$ como anteriormente.

★ 5.17. Propriedades extremais dos valores próprios de uma transformação simétrica

Vamos agora demonstrar que os valores extremos de uma forma quadrática na esfera unitária são valores próprios.

TEOREMA 5.14. *Seja $T: V \rightarrow V$ uma transformação linear simétrica num espaço euclidiano real V , e seja $Q(x) = (T(x), x)$. Entre todos os valores que Q toma na esfera unidade, admita-se que existe um extremo† (máximo ou mínimo) num ponto u com $(u, u) = 1$. Então u é um vector próprio de T ; o correspondente valor próprio é $Q(u)$, o valor extremo de Q na esfera unitária.*

Demonstração. Admitamos que Q tem um mínimo em u . Então temos

$$(5.14) \quad Q(x) \geq Q(u) \quad \text{para todo } x \text{ que verifique } (x, x) = 1.$$

Seja $\lambda = Q(u)$. Se $(x, x) = 1$ temos $Q(u) = \lambda(x, x) = (\lambda x, x)$ pelo que a desigualdade (5.14) pode escrever-se

$$(5.15) \quad (T(x), x) \geq (\lambda x, x)$$

com tanto que $(x, x) = 1$. Demonstramos agora que (5.15) é válida para *qualquer* x em V . Admitamos que $\|x\| = a$. Então $x = ay$, onde $\|y\| = 1$. Por conseguinte

$$(T(x), x) = (T(ay), ay) = a^2(T(y), y) \quad \text{e} \quad (\lambda x, x) = a^2(\lambda y, y).$$

Mas $(T(y), y) \geq (\lambda y, y)$ visto que $(y, y) = 1$. Multiplicando ambos os membros desta igualdade por a^2 obtemos (5.15) para $x = ay$.

Uma vez que $(T(x), x) - (\lambda x, x) = (T(x) - \lambda x, x)$, podemos escrever de novo a desigualdade (5.15) na forma $(T(x) - \lambda x, x) \geq 0$, ou

$$(5.16) \quad (S(x), x) \geq 0, \quad \text{onde} \quad S = T - \lambda I.$$

Quando $x = u$ em (5.14) tem-se uma desigualdade, o mesmo acontecendo em (5.16). A transformação linear S é simétrica. A desigualdade (5.16) estabelece que a forma quadrática Q_1 dada por $Q_1(x) = (S(x), x)$ é não negativa em V . Quando $x = u$ temos

† Se V tem dimensão infinita, a forma quadrática Q não terá necessariamente um extremo na esfera unitária. Será o caso quando T não admite valores próprios. No caso de dimensão finita, Q tem sempre um máximo e um mínimo em pontos da esfera unidade. Isto é uma consequência dum teorema mais geral sobre valores extremos de funções contínuas. Na secção 9.16 apresenta-se um caso particular desse teorema.

$Q_1(u) = 0$. Portanto, pelo teorema 5.13 devemos ter $S(u) = 0$. Por outras palavras, $T(u) = \lambda u$, pelo que u é um vector próprio para T , e $\lambda = Q(u)$ é o correspondente valor próprio. Está assim completada a demonstração se Q tem um mínimo em u .

Se existir um máximo em u todas as desigualdades da demonstração anterior se invertem e aplicamos o teorema 5.13 à forma quadrática não positiva Q_1 .

★ 5.18. O caso de dimensão finita

Suponhamos agora que $\dim V = n$. Então T tem n valores próprios reais os quais podem dispor-se por ordem crescente, por exemplo

$$\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n.$$

Segundo o teorema 5.14, o menor valor próprio λ_1 é o mínimo de Q na esfera unidade, e o maior valor próprio é o máximo de Q na referida esfera. Pretendemos agora demonstrar que os valores próprios intermédios também se apresentam como valores extremos de Q , restringidos a determinados subconjuntos da esfera unidade.

Seja u_1 um vector próprio na esfera unidade que minimiza Q . Então $\lambda_1 = Q(u_1)$. Se λ é um valor próprio de λ_1 qualquer vector próprio correspondente a λ deve ser ortogonal a u_1 . Por conseguinte é natural procurar um tal vector próprio no complemento ortogonal do espaço gerado por u_1 .

Seja S o sub-espaço gerado por u_1 . O complemento ortogonal S^\perp consiste de todos os elementos de V ortogonais a u_1 . Em particular, S^\perp contém de todos os vectores próprios correspondentes aos valores próprios $\lambda \neq \lambda_1$. Com facilidade se verifica que $\dim S^\perp = n - 1$ e que T aplica S^\perp em si próprio.† Seja S_{n-1} a esfera unidade no sub-espaço S^\perp , de dimensão $n - 1$. (A esfera unidade S_{n-1} é um subconjunto da esfera unidade de V). Aplicando o teorema 5.14 ao subespaço S^\perp verificamos que $\lambda_2 = Q(u_2)$, onde u_2 é um ponto que minimiza Q em S_{n-1} .

O vector próprio seguinte, λ_3 , pode obter-se de maneira análoga como o valor mínimo de Q na esfera unidade S_{n-2} no espaço de dimensão $(n - 2)$ formado por todos os elementos ortogonais quer a u_1 quer a u_2 . Continuando este processo, verificamos que cada valor próprio λ_k é o mínimo valor que Q_k toma numa esfera unidade S_{n-k+1} num subespaço de dimensão $n - k + 1$. O maior destes valores mínimos λ_n , é também o valor máximo que Q toma em cada uma das esferas S_{n-k+1} . O correspondente conjunto de vectores próprios u_1, \dots, u_n formam uma base ortonormada para V .

5.19. Transformações unitárias

Concluimos este capítulo com uma breve discussão sobre outra classe importante de transformações conhecidas por transformações unitárias. Na hipótese de dimensão finita são representadas por matrizes unitárias.

† Fez-se isto na demonstração do teorema 5.4 da Secção 5.6.

DEFINIÇÃO. Seja E um espaço euclidiano e V um subespaço de E . Uma transformação linear $T: V \rightarrow E$ diz-se unitária em V se

$$(5.17) \quad (T(x), T(y)) = (x, y) \quad \text{quaisquer que sejam } x \text{ e } y \text{ em } V.$$

Quando E é um espaço euclidiano real uma transformação unitária diz-se também uma transformação ortogonal.

A equação (5.17) significa que T preserva os produtos internos. Deste modo é natural esperar que T conserve também a ortogonalidade e as normas, uma vez que estas derivam do produto interno.

TEOREMA 5.15. Se $T: V \rightarrow E$ é uma transformação unitária em V , então para quaisquer x e y de V tem-se:

- (a) $(x, y) = 0$ implica $(T(x), T(y)) = 0$ (T preserva a ortogonalidade).
- (b) $\|T(x)\| = \|x\|$ (T preserva a norma).
- (c) $\|T(x) - T(y)\| = \|x - y\|$ (T preserva as distâncias).
- (d) T é invertível, e T^{-1} é unitária em $T(V)$.

Demonstração. A alínea (a) é consequência imediata da equação (5.17). A alínea (b) resulta de se fazer $x = y$ em (5.17). A alínea (c) resulta de (b), devido a ser $T(x) - T(y) = T(x - y)$.

Para demonstrar (d) usamos (b) a qual nos mostra que $T(x) = 0$ implica $x = 0$, pelo que T é invertível. Se $x \in T(V)$ e $y \in T(V)$ podemos escrever $x = T(u)$, $y = T(v)$, pelo que se tem

$$(T^{-1}(x), T^{-1}(y)) = (u, v) = (T(u), T(v)) = (x, y).$$

Portanto T^{-1} é unitária em $T(V)$.

Pelo que respeita os valores próprios e vectores próprios temos o seguinte teorema.

TEOREMA 5.16. Seja $T: V \rightarrow E$ uma transformação unitária em V .

- (a) Se T tem um valor próprio λ então $|\lambda| = 1$.
- (b) Se x e y são vectores próprios correspondentes aos valores próprios distintos λ e μ , então x e y são ortogonais.
- (c) Se $V = E$ e $\dim V = n$, e se V é um espaço complexo, então existem vectores próprios u_1, \dots, u_n de T os quais constituem uma base ortonormada para V . A matriz de T relativa a esta base é a matriz diagonal $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, onde λ_k é o valor próprio correspondente a u_k .

Demonstração. Para demonstrar (a), designe x um vector próprio correspondente a λ . Então $x \neq 0$ e $T(x) = \lambda x$. Tomando $y = x$ na equação (5.17) obtemos

$$(\lambda x, \lambda x) = (x, x) \quad \text{ou} \quad \lambda \bar{\lambda} (x, x) = (x, x).$$

Uma vez que $(x, x) > 0$ e $\lambda\bar{\lambda} = |\lambda|^2$, isto implica $|\lambda| = 1$.

Para demonstrar (b), escrevemos $T(x) = \lambda x$, $T(y) = \mu y$ e efectuamos o produto interno $(T(x), T(y))$ de duas maneiras. Temos

$$(T(x), T(y)) = (x, y)$$

visto que T é unitária. Temos também

$$(T(x), T(y)) = (\lambda x, \mu y) = \lambda\bar{\mu}(x, y)$$

já que x e y são vectores próprios. Desta maneira $\lambda\bar{\mu}(x, y) = (x, y)$, pelo que $(x, y) = 0$ a menos que $\lambda\bar{\mu} = 1$. Mas $\lambda\bar{\lambda} = 1$ devido a (a), pelo que se $\lambda\bar{\mu} = 1$ deverá verificar-se $\lambda\bar{\lambda} = \lambda\bar{\mu}$, $\lambda = \bar{\mu}$, $\lambda = \mu$, o que contradiz a afirmação de que λ e μ são distintos. Portanto $\lambda\bar{\mu} \neq 1$ e $(x, y) = 0$.

A alínea (c) demonstra-se por indução em n , de forma muito semelhante à que utilizámos para demonstrar o teorema 5.4 em que se estabelece um resultado análogo para os operadores hermiticos. A única transformação exigida diz respeito à parte da demonstração em que se prova que T aplica S em si próprio, sendo

$$S^\perp = \{x \mid x \in V, (x, u_1) = 0\}.$$

Aqui u_1 é um vector próprio de T com valor próprio λ_1 . Da equação $T(u_1) = \lambda_1 u_1$ deduzimos

$$u_1 = \lambda_1^{-1} T(u_1) = \bar{\lambda}_1 T(u_1)$$

visto que $\lambda_1 \bar{\lambda}_1 = |\lambda_1|^2 = 1$. Escolhamos agora qualquer x em S^\perp e observe-se que

$$(T(x), u_1) = (T(x), \bar{\lambda}_1 T(u_1)) = \lambda_1 (T(x), T(u_1)) = \lambda_1 (x, u_1) = 0.$$

Por conseguinte $T(x) \in S^\perp$ se $x \in S^\perp$ pelo que T aplica S^\perp em si próprio. O resto da demonstração é idêntica à do teorema 5.4, o que nos permite suprimir aqui os seus pormenores.

Os dois teoremas que se seguem dizem respeito a propriedades das transformações unitárias num espaço de dimensão finita. Apenas apresentamos um esboço da demonstração.

TEOREMA 5.17. *Seja $\dim V = n$ e $E = (e_1, \dots, e_n)$ uma base dada em V . Uma transformação linear $T: V \rightarrow V$ é unitária se e só se*

$$(5.18) \quad (T(e_i), T(e_j)) = (e_i, e_j) \quad \text{para todo } i \text{ e } j.$$

Em particular, se E é ortonormada então T é unitária se e só se T aplica E sobre uma base ortonormada.

Indicações para a demonstração. Escreva-se $x = \sum x_i e_i$, $y = \sum y_j e_j$. Então tem-se

$$(x, y) = \left(\sum_{i=1}^n x_i e_i, \sum_{j=1}^n y_j e_j \right) = \sum_{i=1}^n \sum_{j=1}^n x_i y_j (e_i, e_j),$$

e

$$(T(x), T(y)) = \left(\sum_{i=1}^n x_i T(e_i), \sum_{j=1}^n y_j T(e_j) \right) = \sum_{i=1}^n \sum_{j=1}^n x_i y_j (T(e_i), T(e_j)).$$

Compare-se agora (x, y) com $(T(x), T(y))$.

TEOREMA 5.18. Se $\dim V = n$ e (e_1, \dots, e_n) define uma base ortonormada para V e $A = (a_{ij})$ é a representação matricial de uma transformação linear $T: V \rightarrow V$ relativa àquela base, então T é unitária se e só se A é unitária, isto é, se e só se

$$(5.19) \quad A^* A = I.$$

Sugestões para a demonstração. Porque (e_i, e_j) é o elemento ij da matriz identidade, a equação (5.19) implica

$$(5.20) \quad (e_i, e_j) = \sum_{k=1}^n \bar{a}_{ki} a_{kj} = \sum_{k=1}^n a_{ki} \bar{a}_{kj}.$$

Visto ser A a matriz de T temos $T(e_i) = \sum_{k=1}^n a_{ki} e_k$, $T(e_j) = \sum_{r=1}^n a_{rj} e_r$, pelo que

$$(T(e_i), T(e_j)) = \left(\sum_{k=1}^n a_{ki} e_k, \sum_{r=1}^n a_{rj} e_r \right) = \sum_{k=1}^n \sum_{r=1}^n a_{ki} \bar{a}_{rj} (e_k, e_r) = \sum_{k=1}^n a_{ki} \bar{a}_{kj}.$$

Compare-se agora isto com (5.20) e aplique-se o teorema 5.17.

TEOREMA 5.19. Toda a matriz unitária A goza das seguintes propriedades.

- A é não singular e $A^{-1} = A^*$.
- As matrizes A^t , A e A^* são unitárias.
- Os valores próprios de A são números complexos de valor absoluto 1.
- $|\det A| = 1$; se A é real, então $\det A = \pm 1$.

A demonstração do teorema 5.19 é deixada ao leitor como exercício.

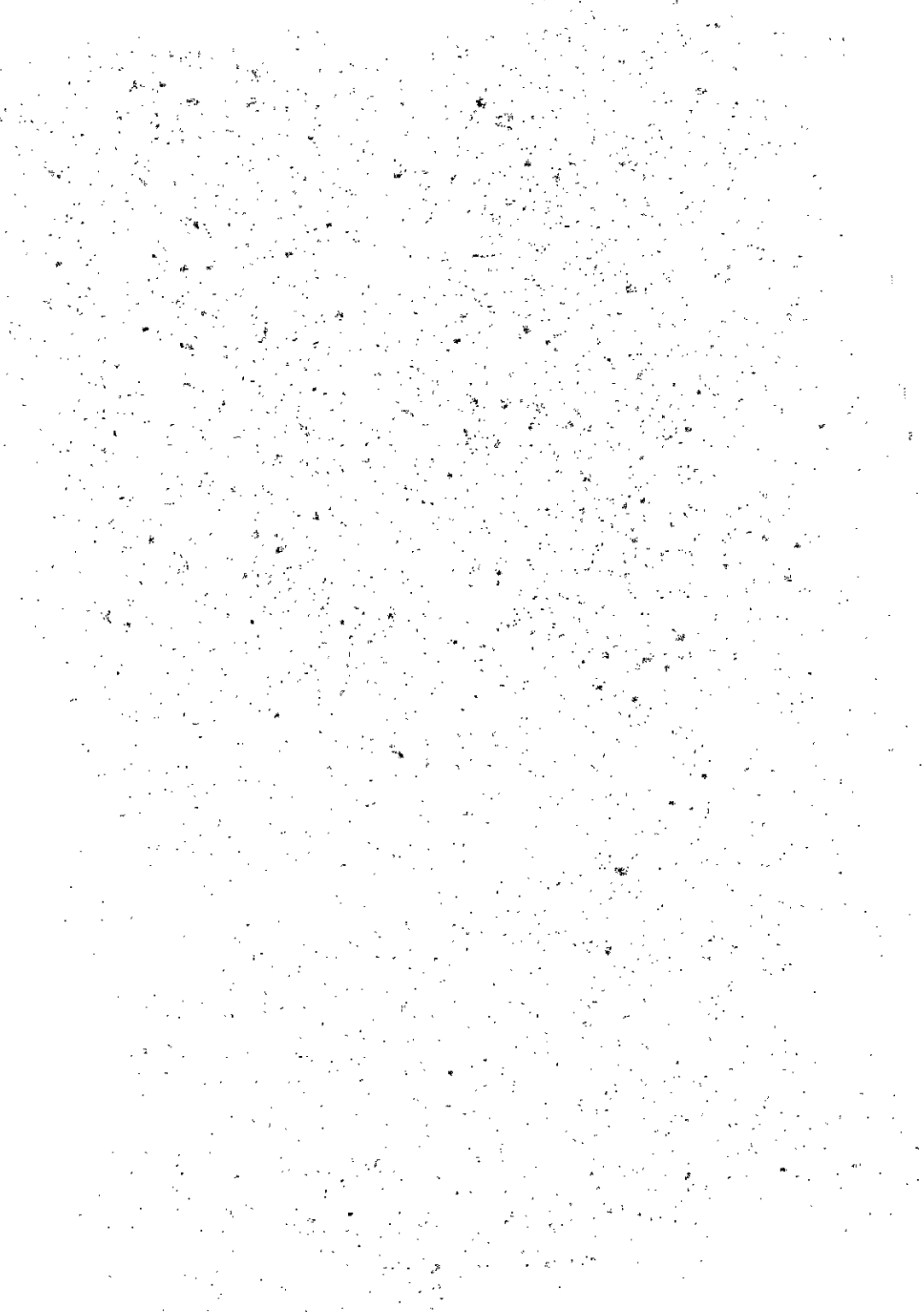
5.20. Exercícios

- (a) Seja $T: V \rightarrow V$ a transformação dada por $T(x) = cx$, onde c é um escalar fixo. Provar que T é unitário se e só se $|c| = 1$.
- (b) Se V é unidimensional, provar que as únicas transformações unitárias em V são as defi-

- nidas em (a). Em particular, se V é um espaço real unidimensional, existem unicamente duas transformações ortogonais, $T(x) = x$ e $T(x) = -x$.
2. Provar cada uma das proposições seguintes relativas à matriz real ortogonal $n \times n$, A .
- (a) Se λ é um valor próprio real de A , então $\lambda = 1$ e $\lambda = -1$.
- (b) Se λ é um valor próprio complexo de A , então o complexo conjugado $\bar{\lambda}$ é também um valor próprio de A . Por outras palavras, os valores próprios não reais de A ocorrem em pares conjugados.
- (c) Se n é ímpar, então A tem pelo menos um valor próprio real.
3. Seja V um espaço real euclidiano de dimensão n . Uma transformação ortogonal $T: V \rightarrow V$ com determinante 1 diz-se uma *rotação*. Se n é ímpar, provar que 1 é um valor próprio para T . Isto prova que cada rotação num espaço de dimensão ímpar tem um eixo fixo [Sugestão: usar o Exercício 2].
4. Dada uma matriz real ortogonal A com -1 como valor próprio de grau de multiplicidade k , provar que $\det A = (-1)^k$.
5. Se T é linear e preserva a norma, provar que T é unitária.
6. Se $T: V \rightarrow V$ é unitária e hermitica, provar que $T^2 = I$.
7. Sejam (e_1, \dots, e_n) e (u_1, \dots, u_n) duas bases ortonormadas para um espaço euclidiano V . Provar que existe uma transformação unitária T a qual aplica uma destas bases na outra.
8. Determinar o real a tal que a matriz seguinte seja unitária

$$\begin{bmatrix} a & \frac{1}{2}i & \frac{1}{2}a(2i-1) \\ ia & \frac{1}{2}(1+i) & \frac{1}{2}a(1-i) \\ a & -\frac{1}{2} & \frac{1}{2}a(2-i) \end{bmatrix}.$$

9. Se A é uma matriz hemi-hermitica, provar que $I - A$ e $I + A$ são não singulares e $(I - A)(I + A)^{-1}$ é unitária.
10. Se A é uma matriz unitária e se $I + A$ é não singular, provar que $(I - A)(I + A)^{-1}$ é hemi-hermitica.
11. Se A é hermitica, provar que $A - iI$ é não singular e que $(A - iI)^{-1}(A + iI)$ é unitária.
12. Provar que qualquer matriz unitária pode ser diagonalizada por uma matriz unitária.
13. Uma matriz quadrada diz-se *normal* se $AA^* = A^*A$. Dizer quais das seguintes matrizes são normais.
- (a) Matrizes hermiticas. (d) Matrizes hemi-simétricas.
- (b) Matrizes hemi-hermiticas. (e) Matrizes unitárias.
- (c) Matrizes simétricas. (f) Matrizes ortogonais.
14. Se A é uma matriz normal ($AA^* = A^*A$) e se U é uma matriz unitária, provar que U^*AU é normal.



6

EQUAÇÕES DIFERENCIAIS LINEARES

6.1. Introdução histórica

A história das equações diferenciais começa no século XVII quando Newton, Leibniz e os Bernoulli resolveram alguns exemplos simples de equações diferenciais de primeira e segunda ordens, postas por alguns problemas de geometria e mecânica. Estes primeiros descobrimentos, iniciados cerca de 1690, pareciam sugerir que as soluções de todas as equações diferenciais, originadas por problemas geométricos e físicos, poderiam ser expressas por intermédio de funções elementares do cálculo. Deste modo, a maior parte do primitivo trabalho foi orientado para o desenvolvimento de técnicas mais ou menos engenhosas tendentes a resolver equações diferenciais por processos elementares como por exemplo a adição, subtração, multiplicação, divisão, composição e integração, aplicadas unicamente um número finito de vezes às funções usuais do cálculo.

Métodos especiais, tais como a separação de variáveis e o uso de factores integrantes foram inventados, de maneira mais ou menos casual, antes do final do século XVII. Durante o século XVIII foram desenvolvidos métodos mais sistematizados, principalmente por Euler, Lagrange e Laplace. Tornou-se de imediato evidente que relativamente poucas equações diferenciais poderiam ser resolvidas por processos elementares. Pouco a pouco os matemáticos começaram a dar-se conta de que era vão o empenho de tentar descobrir métodos gerais para resolver todas as equações diferenciais. Em contrapartida, verificaram ser mais proveitoso averiguar se se sim ou não uma aplicação diferencial dada tem solução e, caso afirmativo, tentar deduzir propriedades da solução a partir da própria equação diferencial. Foi devido a este ponto de vista que os matemáticos começaram a considerar as equações diferenciais como fontes de novas funções.

Uma fase importante da teoria desenvolveu-se nos princípios do século XIX, paralelamente à tendência de conseguir um desenvolvimento melhor estruturado e mais

rigoroso do Cálculo. Em 1820, Cauchy obteve o primeiro "teorema de existência" para as equações diferenciais; tendo provado que toda a equação diferencial de primeira ordem da forma

$$y' = f(x, y)$$

tem uma solução sempre que o segundo membro, $f(x, y)$, satisfaz a certas condições gerais. Um exemplo importante é a equação de Ricatti

$$y' = P(x)y^2 + Q(x)y + R(x),$$

onde P , Q e R são funções dadas. O trabalho de Cauchy garante a existência de uma solução da equação de Ricatti, em qualquer intervalo aberto $(-r, r)$ centrado na origem, desde que P , Q e R admitam desenvolvimentos em série de potências em $(-r, r)$. Em 1841 Joseph Liouville (1809-1882) mostrou que em alguns casos essa solução não pode obter-se por meios elementares.

A experiência tem provado que é difícil obter resultados de grande generalidade relativos às soluções das equações diferenciais, salvo para uns poucos tipos. Entre estes estão as chamadas equações diferenciais *lineares* que se apresentam em grande diversidade de problemas científicos. Alguns tipos simples destas equações foram estudados no Volume I, a saber as de primeira e as de segunda ordem com coeficientes constantes. A secção seguinte é dedicada a uma revisão dos principais resultados obtidos a referentes a essas equações.

6.2. Revisão dos resultados já estabelecidos relativos às equações diferenciais lineares de primeira e de segunda ordem

Uma equação diferencial linear de primeira ordem é da forma

$$(6.1) \quad y' + P(x)y = Q(x),$$

com P e Q funções conhecidas. No Volume I demonstrámos um teorema de existência e unicidade para esta equação (teorema 8.3) que voltamos a enunciar agora.

TEOREMA 6.1. *Sejam P e Q funções contínuas num intervalo aberto J . Seja a um ponto qualquer de J e b um número real qualquer. Então existe uma e uma só função $y = f(x)$ que satisfaz à equação diferencial (6.1) e à condição inicial $f(a) = b$. Esta função é dada pela fórmula explícita*

$$(6.2) \quad f(x) = be^{-A(x)} + e^{-A(x)} \int_a^x Q(t)e^{A(t)} dt,$$

com $A(x) = \int_a^x P(t) dt$.

As equações lineares de segunda ordem são da forma

$$P_0(x)y'' + P_1(x)y' + P_2(x)y = R(x).$$

Se os coeficientes P_0 , P_1 , P_2 e o segundo membro R são contínuas em algum intervalo J e se P_0 não se anula em J , um teorema de existência (exposto na Secção 6.5) garante que sempre existem soluções no intervalo J . Contudo, não existe uma fórmula geral análoga a (6.2) capaz de exprimir essas soluções em função de P_0 , P_1 , P_2 e R . Assim, nesta generalização relativamente simples de (6.1), a teoria está longe de ser completa, excepto em casos particulares. Se os coeficientes são *constantes* e se R é nula, todas as soluções podem determinar-se explicitamente por meio de polinómios e funções exponenciais e trigonométricas como se afirma no teorema seguinte que já foi demonstrado no volume I (teorema 8.7).

TEOREMA 6.2. *Consideremos a equação diferencial*

$$(6.3) \quad y'' + ay' + by = 0,$$

onde a e b são constantes reais dadas. Seja $d = a^2 - 4b$. Toda a solução de (6.3) no intervalo $(-\infty, +\infty)$ tem a forma

$$(6.4) \quad y = e^{-ax/2} [c_1 u_1(x) + c_2 u_2(x)],$$

com c_1 e c_2 constantes, e as funções $u_1(x)$ e $u_2(x)$ determinados de acordo com o sinal de d , conforme se indica:

(a) Se $d = 0$, então $u_1(x) = 1$ e $u_2(x) = x$

(b) Se $d > 0$, então $u_1(x) = e^{kx}$ e $u_2(x) = e^{-kx}$, sendo $k = \frac{1}{2}\sqrt{d}$.

(c) Se $d < 0$, então $u_1(x) = \cos kx$ e $u_2(x) = \sin kx$, sendo $k = \frac{1}{2}\sqrt{-d}$.

O número $d = a^2 - 4b$ é o *discriminante* da equação do segundo grau

$$(6.5) \quad r^2 + ar + b = 0.$$

Esta é a *equação característica* da equação diferencial (6.3). As suas raízes são dados por

$$r_1 = \frac{-a + \sqrt{d}}{2}, \quad r_2 = \frac{-a - \sqrt{d}}{2}.$$

O sinal de d determina a natureza destas raízes. Se $d > 0$ ambas as raízes são reais e a solução em (6.4) pode exprimir-se na forma

$$y = c_1 e^{r_1 x} + c_2 e^{r_2 x}.$$

Se $d < 0$, as raízes r_1 e r_2 são complexos conjugados. Cada uma das funções exponenciais complexas $f_1(x) = e^{r_1 x}$ e $f_2(x) = e^{r_2 x}$ é uma solução complexa da equação diferencial (6.3). Obtemos soluções reais examinando as partes real e imaginária de f_1 e f_2 . Escrevendo $r_1 = -\frac{1}{2}a + ik$, $r_2 = -\frac{1}{2}a - ik$, com $k = \frac{1}{2}\sqrt{-d}$, temos

$$f_1(x) = e^{r_1 x} = e^{-ax/2} e^{ikx} = e^{-ax/2} \cos kx + ie^{-ax/2} \operatorname{sen} kx$$

$$f_2(x) = e^{r_2 x} = e^{-ax/2} e^{-ikx} = e^{-ax/2} \cos kx - ie^{-ax/2} \operatorname{sen} kx.$$

A solução geral que se apresenta em (6.4) é uma combinação linear das partes reais e partes imaginárias de $f_1(x)$ e $f_2(x)$.

3. Exercícios

Os exercícios que se seguem foram seleccionados do Capítulo 8 do Volume I e constituem uma revisão referente às equações diferenciais lineares de primeira e segunda ordem.

Equações diferenciais lineares de primeira ordem. Nos Exercícios 1, 2, 3 resolver o problema com os valores iniciais no intervalo indicado.

1. $y' - 3y = e^{2x}$ em $(-\infty, +\infty)$, com $y = 0$ quando $x = 0$.
2. $xy' - 2y = x^3$ em $(0, +\infty)$, com $y = 1$ quando $x = 1$.
3. $y' + y \operatorname{tg} x = \operatorname{sen} 2x$ em $(-\frac{\pi}{2}, \frac{\pi}{2})$, com $y = 2$ quando $x = 0$.
4. Se uma cultura de bactérias cresce proporcionalmente à quantidade existente em cada instante e se a população duplica ao fim de uma hora, quanto crescerá ao fim de duas horas?
5. Uma curva de equação $y = f(x)$ passa pela origem. Rectas paralelas aos eixos, traçadas a partir de um ponto da curva, formam um rectângulo com os referidos eixos. A curva divide um tal rectângulo em duas partes A e B , uma das quais tem n vezes a área da outra. Determinar a função f .
6. (a) Seja u uma solução, não nula, da equação de segunda ordem $y'' + P(x)y' + Q(x)y = 0$. Provar que a substituição $y = uv$ transforma a equação

$$y'' + P(x)y' + Q(x)y = R(x)$$

numa equação linear de primeira ordem em v .

(b) Obter uma solução não nula da equação $y'' - 4y' + x^2(y' - 4y) = 0$, por simples análise da equação e utilizar o método da alínea (a) para determinar uma solução de

$$y'' - 4y' + x^2(y' - 4y) = 2xe^{-x^2/3}$$

tal que $y = 0$ e $y' = 4$ quando $x = 0$.

Equações diferenciais lineares de segunda ordem com coeficientes constantes. Em cada um dos exercícios 7 a 10, determinar todas as soluções em $(-\infty, +\infty)$.

- | | |
|---------------------|---------------------------|
| 7. $y'' - 4y = 0$. | 9. $y'' - 2y' + 5y = 0$. |
| 8. $y'' + 4y = 0$. | 10. $y'' + 2y' + y = 0$. |

1. Determinar todos os valores da constante k , tais que a equação diferencial $y'' + ky = 0$ admite uma solução não trivial $y = f_k(x)$ para a qual $f_k(0) = f_k(1) = 0$. Para cada um desses valores de k , determinar a correspondente solução $y = f_k(x)$. Considerar os valores de k positivos e negativos.
2. Se (a, b) é um ponto dado no plano e se m é um número real dado, provar que a equação diferencial $y'' + k^2y = 0$ tem precisamente uma solução cujo gráfico passa por (a, b) e tem aí declive m . Discutir separadamente o caso $k = 0$.

13. Para cada alínea, determinar uma equação diferencial linear de segunda ordem verificada por u_1 e u_2
- (a) $u_1(x) = e^x$, $u_2(x) = e^{-x}$.
 - (b) $u_1(x) = e^{2x}$, $u_2(x) = xe^{2x}$.
 - (c) $u_1(x) = e^{-x/2} \cos x$, $u_2(x) = e^{-x/2} \sin x$.
 - (d) $u_1(x) = \sin(2x + 1)$, $u_2(x) = \sin(2x + 2)$.
 - (e) $u_1(x) = \cosh x$, $u_2(x) = \sinh x$.
14. Uma partícula está animada de movimento harmônico simples. Inicialmente o seu deslocamento é 1, a sua velocidade é 1 e a sua aceleração é -12 . Calcular o seu deslocamento e aceleração quando a velocidade é $\sqrt{8}$.

6.4. Equações diferenciais lineares de ordem n

Uma equação diferencial linear de ordem n é da forma

$$(6.6) \quad P_0(x)y^{(n)} + P_1(x)y^{(n-1)} + \cdots + P_n(x)y = R(x).$$

As funções P_0, P_1, \dots, P_n que multiplicam as derivadas das várias ordens da função desconhecida y dizem-se os *coeficientes* da equação. No nosso estudo da equação linear supor-se-á sempre que todos os coeficientes são funções contínuas num certo intervalo J . A palavra "intervalo" designará quer um intervalo limitado, quer ilimitado.

Na equação diferencial (6.6) o coeficiente P_0 desempenha um papel especial, uma vez que determina a ordem da equação. Pontos para os quais $P_0(x) = 0$ chamam-se *pontos singulares* da equação. A presença de pontos singulares introduz, algumas vezes, dificuldades que requerem um estudo especial. Para evitar essas dificuldades admitimos que a função P_0 nunca se anula em J . Então podemos dividir ambos os membros de (6.6) por P_0 e escrever a equação diferencial com o primeiro coeficiente igual a 1. Portanto, no estudo geral supomos sempre que a equação diferencial é da forma

$$(6.7) \quad y^{(n)} + P_1(x)y^{(n-1)} + \cdots + P_n(x)y = R(x).$$

A discussão das equações lineares pode simplificar-se mediante o recurso à notação de operadores. Seja $\mathcal{C}(J)$ o espaço linear de todas as funções de valores reais, contínuas num intervalo J e seja $\mathcal{C}^n(J)$ o subespaço de todas as funções f cujas n primeiras derivadas $f', f'', \dots, f^{(n)}$ existem e são contínuas em J . Seja P_1, \dots, P_n n funções dadas em $\mathcal{C}(J)$ e consideremos o operador $L: \mathcal{C}^n(J) \rightarrow \mathcal{C}(J)$ definido por

$$L(f) = f^{(n)} + P_1 f^{(n-1)} + \cdots + P_n f.$$

O operador L escreve-se, por vezes,

$$L = D^n + P_1 D^{n-1} + \cdots + P_n,$$

com D^k o operador derivação de ordem k . Na notação operacional a equação diferencial (6.7) escreve-se muito simplesmente

$$(6.8) \quad L(y) = R.$$

Uma solução desta equação é qualquer função y em $\mathcal{C}^n(J)$ que verifique (6.8) no intervalo J .

É fácil verificar que $L(y_1 + y_2) = L(y_1) + L(y_2)$ e que $L(cy) = cL(y)$, para qualquer constante c . Quer isto dizer que L é um operador *linear*. Por isso se designa a equação $L(y) = R$ por equação linear. O operador L chama-se *um operador diferencial linear de ordem n* .

A cada equação linear $L(y) = R$ podemos assim associar a equação

$$L(y) = 0,$$

na qual o segundo membro aparece substituído por zero. Esta diz-se a *equação homogénea* correspondente a $L(y) = R$. Quando R não identicamente nulo, a equação $L(y) = R$ diz-se uma *equação não homogénea*. Verificaremos adiante que podemos resolver a equação não homogénea sempre que seja possível resolver a correspondente equação homogénea. Por tal facto iniciaremos o nosso estudo pela equação homogénea.

O conjunto das soluções de uma equação homogénea é o espaço nulo $N(L)$ do operador L . Este chama-se também o *espaço solução* da equação. O espaço solução é um subespaço de $\mathcal{C}^n(J)$. Embora $\mathcal{C}^n(J)$ seja de dimensão infinita, resulta que o espaço solução $N(L)$ tem sempre dimensão finita. Com efeito provaremos que

$$(6.9) \quad \dim N(L) = n,$$

com n a ordem do operador L . A equação (6.9) chama-se o *teorema de dimensionalidade* para operadores diferenciais lineares. O teorema de dimensionalidade estabelecer-se-á como uma consequência de um teorema de existência e unicidade que vamos estudar a seguir.

6.5. O teorema de existência e unicidade

TEOREMA 6.3. TEOREMA DE EXISTÊNCIA E UNICIDADE PARA EQUAÇÕES LINEARES DE ORDEM n . *Sejam P_1, P_2, \dots, P_n funções contínuas num intervalo aberto J , e seja L o operador diferencial linear*

$$L = D^n + P_1 D^{n-1} + \dots + P_n.$$

Se $x_0 \in J$ e se k_0, k_1, \dots, k_{n-1} são n números reais dados, existe uma e uma só função $y = f(x)$ que satisfaz à equação diferencial homogénea $L(y) = 0$ em J e que igualmente satisfaz às condições iniciais

$$f(x_0) = k_0, f'(x_0) = k_1, \dots, f^{(n-1)}(x_0) = k_{n-1}.$$

Nota: O vector do espaço n dimensional definido por $(f(x_0), f'(x_0), \dots, f^{(n-1)}(x_0))$ chama-se o *vector valor inicial* de f em x_0 . O teorema 6.3. diz-nos que se escolhemos um ponto x_0 em J e um vector no espaço n dimensional, a equação homogénea $L(y) = 0$ tem precisamente uma solução $y = f(x)$ em J com aquele vector valor inicial em x_0 . Por exemplo, quando $n = 2$ existe precisamente uma solução com o valor prescrito $f(x_0)$ e a derivada prescrita $f'(x_0)$, no ponto dado x_0 .

A demonstração do teorema de existência e unicidade obter-se-á como um corolário dos teoremas de existência e unicidade mais gerais que se estudam do Capítulo 7. Na Secção 7.9 apresenta-se outra demonstração para o mesmo teorema, para o caso de equações de coeficientes constantes.

6.6. A dimensão do espaço solução de uma equação linear homogénea

TEOREMA 6.4. TEOREMA DE DIMENSIONALIDADE. *Seja $L: \mathcal{C}^n(J) \rightarrow \mathcal{C}(J)$ um operador diferencial linear de ordem n , dado por*

$$(6.10) \quad L = D^n + P_1 D^{n-1} + \dots + P_n.$$

Então o espaço solução da equação $L(y) = 0$ tem dimensão n .

Demonstração. Seja V_n o espaço linear n -dimensional dos n -tuplos de escalares. Seja T a transformação linear que aplica cada função f do espaço solução $N(L)$ sobre o vector valor inicial de f em x_0 ,

$$T(f) = (f(x_0), f'(x_0), \dots, f^{(n-1)}(x_0)),$$

onde x_0 é um ponto fixo de J . O teorema de unicidade diz-nos que $T(f) = 0$ implica $f = 0$. Portanto, pelo teorema 2.10, T é biunívoca em $N(L)$. Logo T^{-1} é também biunívoca e aplica V_n sobre $N(L)$, e o teorema 2.11 mostra que $\dim N(L) = \dim V_n = n$.

Uma vez sabido que o espaço solução tem dimensão n , resulta que qualquer conjunto de n soluções independentes servirá como base. Portanto, como corolário do teorema de dimensionalidade temos:

TEOREMA 6.5. *Seja $L: \mathcal{C}^n(J) \rightarrow \mathcal{C}(J)$ um operador diferencial linear de ordem n . Se u_1, \dots, u_n são n soluções independentes da equação diferencial homogénea $L(y) = 0$ em J , então toda a solução $y = f(x)$ em J pode exprimir-se na forma*

$$(6.11) \quad f(x) = \sum_{k=1}^n c_k u_k(x),$$

onde c_1, c_2, \dots, c_k são constantes.

Nota: Uma vez que todas as soluções da equação diferencial $L(y) = 0$ estão contidas na fórmula (6.11) a combinação linear do segundo membro, com constantes arbitrárias c_1, \dots, c_n , chama-se a *solução geral* da equação diferencial.

O teorema de dimensionalidade diz-nos que o espaço solução de uma equação diferencial linear homogénea de ordem n tem sempre uma base de n soluções, mas não nos diz como determinar tal base. Com efeito, não se conhece nenhum método simples para determinar uma base de soluções para toda a equação diferencial linear. Contudo, têm-se inventado métodos especiais para equações particulares. Entre estas estão as equações diferenciais com coeficientes constantes que voltamos a considerar.

6.7. A álgebra de operadores de coeficientes constantes

Um operador A de coeficientes constantes é um operador linear da forma

$$(6.12) \quad A = a_0 D^n + a_1 D^{n-1} + \cdots + a_{n-1} D + a_n,$$

onde D é o operador derivada e $a_0, a_1, \dots, a_{n-1}, a_n$ são constantes reais. Se $a_0 \neq 0$ o operador diz-se de ordem n . O operador A pode aplicar-se a qualquer função y com derivadas até à ordem n em um certo intervalo, sendo o resultado uma função $A(y)$ dada por

$$A(y) = a_0 y^{(n)} + a_1 y^{(n-1)} + \cdots + a_{n-1} y' + a_n y.$$

Nesta Secção interessa-nos considerar funções possuindo derivadas de todas as ordens em $(-\infty, +\infty)$. O conjunto de tais funções representa-se por \mathcal{C}^∞ e será referido como a classe das *funções infinitamente deriváveis*. Se $y \in \mathcal{C}^\infty$ então $A(y)$ pertence também a \mathcal{C}^∞ .

As operações algébricas usuais com transformações lineares (*adição, multiplicação por escalares e composição* ou *multiplicação*) podem aplicar-se, em particular, aos operadores de coeficientes constantes. Sejam A e B dois operadores de coeficientes constantes (não necessariamente da mesma ordem). Uma vez que a soma $A + B$ e os produtos λA são também operadores com coeficientes constantes, o conjunto de todos os operadores de coeficientes constantes é um espaço linear. O produto de A e B (por qualquer ordem) é também um operador de coeficientes constantes. Deste modo somas, produtos, e produtos por escalares, de operadores de coeficientes constantes satisfazem às propriedades comutativa, associativa e distributiva, como quaisquer outras transformações lineares. Igualmente, porque $D^r D^s = D^s D^r$ para quaisquer inteiros positivos r e s dois quaisquer operadores de coeficientes constantes são *permutáveis*: $AB = BA$.

A cada operador de coeficientes constantes A associamos um polinómio p_A , chamado o polinómio característico de A . Se A é dado por (6.12), p_A é o polinómio que tem os mesmos coeficientes que A , isto é, para todo real r tem-se

$$p_A(r) = a_0 r^n + a_1 r^{n-1} + \cdots + a_n.$$

Inversamente, dado um polinómio real qualquer p , existe um correspondente operador A cujos coeficientes são os mesmos de p . O teorema seguinte mostra que esta

associação entre operadores e polinómios é uma correspondência biunívoca. Além disso, esta correspondência associa a somas e produtos de operadores e produtos de operadores por escalares as correspondentes somas e produtos dos respectivos polinómios característicos e produtos destes por escalares.

TEOREMA 6.6. *Se A e B são operadores de coeficientes constantes com polinómios característicos p_A e p_B , respectivamente, e λ um número real, então tem-se:*

(a) $A = B$ se e só se $p_A = p_B$,

(b) $p_{A+B} = p_A + p_B$,

(c) $p_{AB} = p_A \cdot p_B$,

(d) $p_{\lambda A} = \lambda \cdot p_A$.

Demonstração. Consideremos em primeiro lugar, a alínea (a). Suponhamos $p_A = p_B$. Pretendemos provar que $A(y) = B(y)$ para todo y em \mathcal{C}^∞ . Uma vez que $p_A = p_B$, ambos os polinómios têm o mesmo grau e os mesmos coeficientes. Por conseguinte A e B têm a mesma ordem e os mesmos coeficientes, pelo que $A(y) = B(y)$ para cada y em \mathcal{C}^∞ .

Demonstremos agora que $A = B$ implica que $p_A = p_B$. A relação $A = B$ significa que $A(y) = B(y)$ para todo y em \mathcal{C}^∞ . Façamos $y = e^{rx}$, com r constante. Uma vez que $y^{(k)} = r^k e^{rx}$ para cada $k \geq 0$, tem-se

$$A(y) = p_A(r)e^{rx} \quad \text{e} \quad B(y) = p_B(r)e^{rx}.$$

A equação $A(y) = B(y)$ implica $p_A(r) = p_B(r)$. Uma vez que r é arbitrária deve ter-se $p_A = p_B$, estando assim demonstrada a alínea (a).

As alíneas (b), (c) e (d) resultam imediatamente da definição de polinómio característico.

Do teorema 6.6 resulta que toda a relação algébrica que inclua somas, e produtos por escalares dos polinómios p_A e p_B também é válida para os operadores A e B . Em particular, se o polinómio característico p_A pode ser factorizado como um produto de dois ou mais polinómios, cada factor deve ser o polinómio característico de um certo operador com coeficientes constantes, pelo que, pelo teorema 6.6, existe uma factorização correspondente do operador A . Por exemplo, se $p_A(r) = p_B(r)p_C(r)$, então $A = BC$. Se $p_A(r)$ pode factorizar-se como um produto de n factores lineares, por exemplo

$$(6.13) \quad p_A(r) = a_0(r - r_1)(r - r_2) \cdots (r - r_n),$$

a factorização correspondente de A toma a forma

$$A = a_0(D - r_1)(D - r_2) \cdots (D - r_n).$$

O teorema fundamental da álgebra diz-nos que todo o polinómio $p_A(r)$ de grau $n \geq 1$ admite uma factorização da forma (6.13), sendo r_1, r_2, \dots, r_n as raízes da equação

$$p_A(r) = 0,$$

chamada a *equação característica* de A . Cada raiz escreve-se tantas vezes quantas as unidades do seu grau de multiplicidade. As raízes podem ser reais ou complexas: Porque $p_A(r)$ tem coeficientes reais, as raízes complexas aparecem em pares conjugados, $\alpha + i\beta$, $\alpha - i\beta$, se $\beta \neq 0$. Os dois factores lineares, correspondentes a cada tal par de raízes, podem combinar-se para dar um factor quadrático $r^2 - 2\alpha r + \alpha^2 + \beta^2$ cujos coeficientes são reais. Deste modo, todo o polinómio $p_A(r)$ pode factorizar-se num produto de polinómios lineares e quadráticos *com coeficientes reais*. Tal facto proporciona-nos uma factorização correspondente do operador A como um produto de operadores com coeficientes constantes de primeira e segunda ordem com coeficientes reais.

EXEMPLO 1. Seja $A = D^2 - 5D + 6$. Porque o polinómio característico $p_A(r)$ admite a factorização $r^2 - 5r + 6 = (r - 2)(r - 3)$, o operador A admite a factorização

$$D^2 - 5D + 6 = (D - 2)(D - 3).$$

EXEMPLO 2. Seja $A = D^4 - 2D^3 + 2D^2 - 2D + 1$. O polinómio característico $p_A(r)$ admite a factorização

$$r^4 - 2r^3 + 2r^2 - 2r + 1 = (r - 1)(r - 1)(r^2 + 1),$$

pelo que A tem a factorização

$$A = (D - 1)(D - 1)(D^2 + 1).$$

6.8. Determinação de uma base de soluções para equações lineares com coeficientes constantes por factorização de operadores

O teorema que se apresenta a seguir mostra como a factorização de operadores com coeficientes constantes nos auxilia a resolver equações diferenciais lineares com coeficientes constantes.

TEOREMA 6.7. *Se L é um operador com coeficientes constantes que pode ser factorizado como um produto de operadores com coeficientes constantes, por exemplo*

$$L = A_1 A_2 \cdots A_k.$$

então o espaço solução da equação diferencial linear $L(y) = 0$ contém o espaço solução de cada uma das equações diferenciais $A_i(y) = 0$, ou, por outras palavras

$$(6.14) \quad N(A_i) \subseteq N(L) \quad \text{para todo } i = 1, 2, \dots, k.$$

Demonstração. Se u é o espaço nulo do último factor A_n tem-se $A_k(u) = 0$, pelo que

$$L(u) = (A_1 A_2 \cdots A_k)(u) = (A_1 \cdots A_{k-1})A_k(u) = (A_1 \cdots A_{k-1})(0) = 0.$$

Deste modo o espaço nulo de L contém o espaço nulo do último factor A_k . Mas porque os operadores de coeficientes constantes permutam, podem sempre reordenar-se os factores de tal modo que qualquer um deles seja o último e portanto está demonstrada (6.14).

Se $L(u) = 0$, o operador L diz-se *anular* u . O Teorema 6.7 diz-nos que se um factor A_i de L anula u , então L também anula u .

Vamos mostrar como pode o teorema ser utilizado para resolver equações diferenciais homogêneas com coeficientes constantes. Escolheremos exemplos que ilustrem aspectos diferentes, dependentes da natureza das raízes da equação característica.

CASO I. Raízes reais e distintas

EXEMPLO 1. Determinar uma base de soluções para a equação diferencial

$$(6.15) \quad (D^3 - 7D + 6)y = 0.$$

Resolução. A equação é da forma $L(y) = 0$ com

$$L = D^3 - 7D + 6 = (D - 1)(D - 2)(D + 3).$$

O espaço nulo de $D - 1$ contém $u_1(x) = e^x$; o de $D - 2$ contém $u_2(x) = e^{2x}$; e o de $D + 3$ contém $u_3(x) = e^{-3x}$. No capítulo 1 (pg. 11) demonstrou-se que u_1, u_2, u_3 são independentes. Porque três soluções independentes de uma equação de terceira ordem formam uma base para o espaço solução, a solução geral de (6.15) é dada por

$$y = c_1 e^x + c_2 e^{2x} + c_3 e^{-3x}.$$

O método usado para resolver o Exemplo 1 permite-nos encontrar uma base para o espaço solução de qualquer operador com coeficientes constantes que possa decompor-se num produto de factores lineares distintos.

TEOREMA 6.8. Se L for um operador de coeficientes constantes cuja equação característica $p_L(r) = 0$ admite n raízes reais e distintas r_1, r_2, \dots, r_n , então a solução geral da equação diferencial $L(y) = 0$, no intervalo $(-\infty, +\infty)$, é dada pela fórmula

$$(6.16) \quad y = \sum_{k=1}^n c_k e^{r_k x}.$$

Demonstração. Tem-se a factorização

$$L = a_0(D - r_1)(D - r_2) \cdots (D - r_n).$$

Porque o espaço nulo de $(D - r_k)$ contém $u_k(x) = e^{rkx}$, o espaço nulo de L contém as n funções

$$(6.17) \quad u_1(x) = e^{r_1x}, \quad u_2(x) = e^{r_2x}, \dots, u_n(x) = e^{r_nx}.$$

No capítulo 1 (pag. 11) provou-se que estas funções são independentes, portanto formam uma base para o espaço solução da equação $L(y) = 0$, pelo que a solução geral é dada por (6.16).

CASO II. Raízes reais, algumas das quais múltiplas.

Se todas as raízes forem reais, mas não distintas, as funções (6.17) não são independentes e consequentemente não constituem uma base para o espaço solução. Se uma raiz r tem multiplicidade m , então $(D - r)^m$ é um factor de L . O teorema que se segue mostra como obter m soluções independentes no espaço nulo deste factor.

TEOREMA 6.9. As m funções

$$u_1(x) = e^{rx}, \quad u_2(x) = xe^{rx}, \dots, u_m(x) = x^{m-1}e^{rx}$$

são m elementos independentes anulados pelo operador $(D - r)^m$.

Demonstração. A independência destas funções resulta da independência dos polinómios $1, x, x^2, \dots, x^{m-1}$. Para provar que u_1, u_2, \dots, u_m são anuladas por $(D - r)^m$ recorrendo à indução em m .

Se $m = 1$ existe unicamente uma função, $u_1(x) = e^{rx}$, a qual é evidentemente anulada por $(D - r)$. Admitamos, então, que o teorema é verdadeiro para $m - 1$. Tal significa que as funções u_1, \dots, u_{m-1} são anuladas por $(D - r)^{m-1}$. Uma vez que

$$(D - r)^m = (D - r)(D - r)^{m-1}$$

as funções u_1, \dots, u_{m-1} são também anuladas por $(D - r)^m$. Para completar a demonstração torna-se necessário provar que $(D - r)^m$ anula u_m . Posto que

$$(D - r)^m u_m = (D - r)^{m-1} (D - r)(x^{m-1}e^{rx}).$$

tem-se

$$\begin{aligned} (D - r)(x^{m-1}e^{rx}) &= D(x^{m-1}e^{rx}) - rx^{m-1}e^{rx} \\ &= (m-1)x^{m-2}e^{rx} + x^{m-1}re^{rx} - rx^{m-1}e^{rx} \\ &= (m-1)x^{m-2}e^{rx} = (m-1)u_{m-1}(x). \end{aligned}$$

Quando aplicamos $(D - r)^{m-1}$ a ambos os membros desta equação obtemos 0 no segundo membro, uma vez que $(D - r)^{m-1}$ anula u_{m-1} . Por conseguinte $(D - r)^m u_m = 0$, pelo que u_m é anulado por $(D - r)^m$. Está pois completada a demonstração.

EXEMPLO 2. Determinar a solução geral da equação diferencial $L(y) = 0$, onde $L = D^3 - D^2 - 8D + 12$.

Resolução. O operador L admite a factorização

$$L = (D - 2)^2(D + 3).$$

Pelo teorema 6.9, as duas funções

$$u_1(x) = e^{2x}, \quad u_2(x) = xe^{2x}$$

pertencem ao espaço nulo de $(D - 2)^2$. A função $u_3(x) = e^{-3x}$ pertence ao espaço nulo de $(D + 3)$. Uma vez que u_1, u_2, u_3 são independentes (ver Exercício 17 da Secção 6.9) constituem uma base do espaço nulo de L , pelo que a solução geral da equação diferencial é

$$y = c_1 e^{2x} + c_2 x e^{2x} + c_3 e^{-3x}.$$

O teorema 6.9 diz-nos como determinar uma base de soluções para uma equação diferencial linear de ordem n de coeficientes constantes, na hipótese da equação característica possuir apenas raízes reais ainda que algumas delas sejam múltiplas. Se as raízes distintas são, r_1, r_2, \dots, r_k e se aparecem com graus de multiplicidade m_1, m_2, \dots, m_k , a parte da base correspondente a r_p é dada por m_p funções

$$u_{q,p}(x) = x^{q-1} e^{r_p x}, \quad \text{onde } q = 1, 2, \dots, m_p.$$

Quando p toma os valores $1, 2, \dots, k$ obtemos no total $m_1 + \dots + m_k$ funções. No Exercício 17 da Secção 6.9 delineamos uma demonstração para provar que todas estas funções são independentes. Uma vez que a soma das multiplicidades $m_1 + \dots + m_k$ é igual a n , a ordem da equação, as funções $u_{p,q}$ formam uma base para o espaço solução da equação.

EXEMPLO 3. Integrar a equação $(D^6 + 2D^5 - 2D^3 - D^2)y = 0$.

Resolução. Tem-se $D^6 + 2D^5 - 2D^3 - D^2 = D^2(D - 1)(D + 1)^3$. A parte da base correspondente ao factor D^2 é $u_1(x) = 1, u_2(x) = x$; a parte correspondente ao factor $(D - 1)$ é $u_3(x) = e^x$; e a parte correspondente ao factor $(D + 1)^3$ é $u_4(x) = e^{-x}, u_5(x) = x e^{-x}, u_6(x) = x^2 e^{-x}$. As seis funções u_1, \dots, u_6 são independentes, pelo que a solução geral da equação é

$$y = c_1 + c_2 x + c_3 e^x + (c_4 + c_5 x + c_6 x^2) e^{-x}.$$

CASO III. Raízes complexas.

Se se utilizam exponenciais complexas, não há necessidade de distinguir entre raízes reais e complexas na equação característica da equação diferencial $L(y) = 0$. Se se de-

sejam soluções de valor real, factoriza-se o operador L em factores lineares e quadráticos com coeficientes reais. Cada par de raízes complexas conjugadas $\alpha + \beta i$, $\alpha - \beta i$ corresponde a um factor quadrático.

$$(6.18) \quad D^2 - 2\alpha D + \alpha^2 + \beta^2.$$

O espaço nulo deste operador de segunda ordem contém as duas funções independentes $u(x) = e^{\alpha x} \cos \beta x$ e $v(x) = e^{\alpha x} \sin \beta x$. Se o par de raízes $\alpha \pm i\beta$ tem multiplicidade m , então o factor quadrático aparece elevado à potência m . O espaço nulo do operador

$$[D^2 - 2\alpha D + \alpha^2 + \beta^2]^m$$

contém $2m$ funções independentes

$$u_q(x) = x^{q-1} e^{\alpha x} \cos \beta x, \quad v_q(x) = x^{q-1} e^{\alpha x} \sin \beta x, \quad q = 1, 2, \dots, m.$$

A demonstração desta proposição pode efectuar-se facilmente, por indução em m . (No Exercício 20 da Secção 6.9 apresentam-se esboços de demonstrações). Os exemplos seguintes ilustram algumas das hipóteses.

EXEMPLO 4. $y'' - 4y' + 13y = 0$. A equação característica, $r^2 - 4r + 13 = 0$, tem as raízes $0, 2 \pm 3i$; a solução geral é

$$y = c_1 + e^{2x}(c_2 \cos 3x + c_3 \sin 3x).$$

EXEMPLO 5. $y''' - 2y'' + 4y' - 8y = 0$. A equação característica é

$$r^3 - 2r^2 + 4r - 8 = (r - 2)(r^2 + 4) = 0;$$

As suas raízes são $2, 2i, -2i$, pelo que a solução geral da equação diferencial é

$$y = c_1 e^{2x} + c_2 \cos 2x + c_3 \sin 2x.$$

EXEMPLO 6. $y^{(5)} - 9y^{(4)} + 34y''' - 66y'' + 65y' - 25y = 0$. A equação característica pode escrever-se

$$(r - 1)(r^2 - 4r + 5)^2 = 0;$$

as suas raízes são $1, 2 \pm i, 2 \pm i$, de maneira que a solução geral da equação diferencial é

$$y = c_1 e^x + e^{2x}[(c_2 + c_3 x) \cos x + (c_4 + c_5 x) \sin x].$$

6.9. Exercícios

Determinar a solução geral de cada uma das equações diferenciais dos Exercícios 1 a 12

1. $y''' - 2y'' - 3y' = 0$.
2. $y''' - y' = 0$.
3. $y''' + 4y'' + 4y' = 0$.
4. $y''' - 3y'' + 3y' - y = 0$.
5. $y^{(4)} + 4y''' + 6y'' + 4y' + y = 0$.
6. $y^{(4)} - 16y = 0$.
7. $y^{(4)} + 16y = 0$.
8. $y''' - y = 0$.
9. $y^{(4)} + 4y''' + 8y'' + 8y' + 4y = 0$.
10. $y^{(4)} + 2y'' + y = 0$.
11. $y^{(6)} + 4y^{(4)} + 4y'' = 0$.
12. $y^{(6)} + 8y^{(4)} + 16y'' = 0$.

13. Se m é uma constante positiva, determinar a solução particular $y = f(x)$ da equação diferencial

$$y''' - my'' + m^2y' - m^3y = 0$$

que satisfaz às condições $f(0) = f'(0) = 0, f''(0) = 1$.

14. Uma equação diferencial linear de coeficientes constantes tem a equação característica $f(r) = 0$. Se todas as raízes da equação característica forem negativas, provar que toda a solução da equação diferencial tende para zero quando $x \rightarrow +\infty$. O que pode concluir-se acerca do comportamento de todas as soluções no intervalo $[0, +\infty)$ se todas as raízes da equação característica são não positivas?
15. Em cada alínea, determinar uma equação diferencial linear com coeficientes constantes verificada por todas as funções que se indicam
- (a) $u_1(x) = e^x, \quad u_2(x) = e^{-x}, \quad u_3(x) = e^{2x}, \quad u_4(x) = e^{-2x}$.
 - (b) $u_1(x) = e^{-2x}, \quad u_2(x) = xe^{-2x}, \quad u_3(x) = x^2e^{-2x}$.
 - (c) $u_1(x) = 1, \quad u_2(x) = x, \quad u_3(x) = e^x, \quad u_4(x) = xe^x$.
 - (d) $u_1(x) = x, \quad u_2(x) = e^x, \quad u_3(x) = xe^x$.
 - (e) $u_1(x) = x^2, \quad u_2(x) = e^x, \quad u_3(x) = xe^x$.
 - (f) $u_1(x) = e^{-2x} \cos 3x, \quad u_2(x) = e^{-2x} \sin 3x, \quad u_3(x) = e^{-2x}, \quad u_4(x) = xe^{-2x}$.
 - (g) $u_1(x) = \cosh x, \quad u_2(x) = \sinh x, \quad u_3(x) = x \cosh x, \quad u_4(x) = x \sinh x$.
 - (h) $u_1(x) = \cosh x \sinh x, \quad u_2(x) = \sinh x \cosh x, \quad u_3(x) = x$.
16. Seja r_1, \dots, r_n, n números reais distintos e sejam Q_1, \dots, Q_n n polinómios, nenhum dos quais é o polinómio nulo. Provar que as n funções

$$u_1(x) = Q_1(x)e^{r_1x}, \dots, u_n(x) = Q_n(x)e^{r_nx}$$

são independentes.

Esboço da demonstração. Utilizar a indução em n . Para $n = 1$ e $n = 2$ o resultado é facilmente verificável. Suponhamos que a proposição é verdadeira para $n = p$ e sejam $c_1, \dots, c_p, \dots, c_{p+1}, p + 1$ escalares reais tais que

$$\sum_{k=1}^{p+1} c_k Q_k(x) e^{r_k x} = 0.$$

Multiplicar ambos os membros por $e^{-r_{p+1}x}$ e derivar a equação obtida. Usar então a hipótese de indução para mostrar que todos os escalares c_k são 0. Pode dar-se outra demonstração baseada na ordem de grandeza quando $x \rightarrow +\infty$, como foi feito no Exemplo 7 da Secção 1.7 (pg. 11).

17. Sejam m_1, m_2, \dots, m_k , k inteiros positivos e r_1, r_2, \dots, r_k , k números reais distintos e $n = m_1 + \dots + m_k$. Para cada par de inteiros p, q satisfazendo a $1 \leq p \leq k, 1 \leq q \leq m_p$, seja

$$u_{q,p}(x) = x^{q-1} e^{r_p x}.$$

Por exemplo quando $p = 1$ as funções correspondentes são

$$u_{1,1}(x) = e^{r_1 x}, \quad u_{2,1}(x) = x e^{r_1 x}, \dots, u_{m_1,1}(x) = x^{m_1-1} e^{r_1 x}.$$

Provar que as n funções $u_{q,p}$ assim definidas são independentes. [Sugestão: Recorrer ao Exercício 16.]

18. Seja L um operador diferencial linear de coeficientes constantes de ordem n com polinômio característico $p(r)$. Seja L' o operador diferencial linear de coeficientes constantes cujo polinômio característico é o polinômio derivado $p'(r)$. Por exemplo, se $L = 2D^2 - 3D + 1$ então $L' = 4D - 3$. Mais geralmente, defina-se a derivada de ordem m , $L^{(m)}$, como sendo o operador cujo polinômio característico é a derivada $p^{(m)}(r)$. (O operador $L^{(m)}$ não deve confundir-se com a potência L^m).

(a) Se u admite derivadas até à ordem n , provar que

$$L(u) = \sum_{k=0}^n \frac{p^{(k)}(0)}{k!} u^{(k)}.$$

(b) Se u admite derivadas até à ordem $n - m$, provar que

$$L^{(m)}(u) = \sum_{k=0}^{n-m} \frac{p^{(k+m)}(0)}{k!} u^{(k)} \quad \text{para } m = 0, 1, 2, \dots, n, \quad \text{onde}$$

$$L^{(0)} = L.$$

19. Com a notação do Exercício 18, se u e v têm derivadas até à ordem n , provar que

$$L(uv) = \sum_{k=0}^n \frac{L^{(k)}(u)}{k!} v^{(k)}.$$

[Sugestão: Utilizar o Exercício 18, conjuntamente com a fórmula de Leibniz para a derivada de ordem n de um produto:

$$(uv)^{(k)} = \sum_{r=0}^k \binom{k}{r} u^{(k-r)} v^{(r)}.]$$

20. (a) Seja $p(t) = g(t)^{mr}(t)$, com q e p polinômios e m um inteiro positivo. Provar que $p'(t) = q(t)^{m-1} s(t)$, com s um polinômio.
 (b) Seja L um operador diferencial com coeficientes constantes que anula u , onde u é uma função dada de x . Seja $M = L^m$, a potência de L de ordem m , $m > 1$. Provar que cada uma das derivadas $M', M'', \dots, M^{(m-1)}$ também anula u .
 (c) Servir-se da alínea (b) e do Exercício 19 para provar que M anula cada uma das funções $u, xu, \dots, x^{m-1}u$.
 (d) Servir-se da alínea (c) para provar que o operador $(D^2 - 2\alpha D + \alpha^2 + \beta^2)$ anula cada uma das funções $x^q e^{\alpha x} \sin \beta x$ e $x^q e^{\alpha x} \cos \beta x$ para $q = 1, 2, \dots, m-1$.

21. Seja L um operador de ordem n com coeficientes constantes e cujo polinómio característico é $p(r)$. Se α é constante e se u tem n derivadas, provar que

$$L(e^{\alpha x} u(x)) = e^{\alpha x} \sum_{k=0}^n \frac{p^{(k)}(\alpha)}{k!} u^{(k)}(x).$$

6.10. Relação entre as equações homogêneas e não homogêneas

Voltamos de novo à equação diferencial linear de ordem n com coeficientes não necessariamente constantes. O teorema que se segue estabelece a relação entre as soluções da equação homogênea $L(y) = 0$ e as de uma equação não homogênea $L(y) = R(x)$.

TEOREMA 6.10. *Seja $L: \mathcal{C}_n(J) \rightarrow \mathcal{C}(J)$ um operador diferencial linear de ordem n . Sejam u_1, \dots, u_n n soluções independentes da equação homogênea $L(y) = 0$, e y_1 uma solução particular da equação não homogênea $L(y) = R$, onde $R \in \mathcal{C}(J)$. Toda a solução $y = f(x)$ da equação não homogênea é da forma*

$$(6.19) \quad f(x) = y_1(x) + \sum_{k=1}^n c_k u_k(x),$$

onde c_1, \dots, c_n são constantes.

Demonstração. Pela linearidade temos $L(f - y_1) = L(f) - L(y_1) = R - R = 0$. Portanto $f - y_1$ pertence ao espaço solução da equação homogênea $L(y) = 0$, pelo que $f - y_1$ é uma combinação linear de u_1, \dots, u_n , por exemplo $f - y_1 = c_1 u_1 + \dots + c_n u_n$, estando assim demonstrado (6.19).

Uma vez que todas as soluções de $L(y) = R$ estão contidas em (6.19), a soma do segundo membro de (6.19) (com constantes arbitrárias c_1, c_2, \dots, c_n) chama-se a *solução geral* da equação não homogênea. O teorema 6.10 estabelece que a solução geral da equação não homogênea se obtém somando a y_1 a solução geral da equação homogênea.

Nota: O teorema 6.10 admite uma interpretação geométrica simples, a qual nos ajuda a clarificar o seu significado. Para determinar todos os pontos de um plano determinamos um ponto particular do plano e juntamos-lhe todos os pontos de um plano paralelo que passa pela origem. Para determinar todas as soluções de $L(y) = R$, determinamos uma solução particular e somamos a esta todas as soluções da equação homogênea $L(y) = 0$. O conjunto de soluções da equação não homogênea é análogo ao plano que passa por um determinado ponto. O espaço solução da equação homogênea é análogo ao plano paralelo que passa pela origem.

Na prática para utilizar o teorema 6.10 temos que resolver dois problemas: (1) Determinar a solução geral de equação homogênea $L(y) = 0$, e (2) determinar uma solução particular da equação não homogênea $L(y) = R$. Vamos ver em seguida que podemos sempre resolver o problema (2) se podermos resolver o problema (1).

6.11. Determinação de uma solução particular da equação não homogénea. O método de variação das constantes

Voltamos agora a nossa atenção para o problema de determinação de uma solução particular y_1 da equação não homogénea $L(y) = R$. Vamos tratar um método conhecido por *variação das constantes* que nos indica como determinar y_1 se conhecermos n soluções independentes u_1, \dots, u_n da equação homogénea $L(y) = 0$. O método proporciona uma solução particular da forma

$$(6.20) \quad y_1 = v_1 u_1 + \dots + v_n u_n,$$

em que v_1, \dots, v_n são funções que podem ser calculadas por intermédio de u_1, \dots, u_n e do segundo membro R . O método conduz a um sistema de n equações algébricas lineares, satisfeitas pelas derivadas v'_1, \dots, v'_n . Este sistema pode sempre resolver-se porque a respectiva matriz dos coeficientes é não singular. A integração das derivadas dá-nos as funções pretendidas v_1, \dots, v_n . O método foi pela primeira vez usado por Johann Bernoulli para resolver equações lineares de primeira ordem e depois por Lagrange em 1774 para resolver equações lineares de segunda ordem.

Para o caso de uma equação de ordem n os pormenores podem simplificar-se pelo uso de notação vectorial e matricial. O segundo membro de (6.20) pode escrever-se como um produto interno,

$$(6.21) \quad y_1 = (v, u),$$

onde v e u são funções vectoriais n dimensionais dadas por

$$v = (v_1, \dots, v_n), \quad u = (u_1, \dots, u_n).$$

Tentamos escolher v de tal maneira que o produto interno definido y_1 satisfaça à equação não homogénea $L(y) = R$, dado que $L(u) = 0$, com $L(u) = (L(u_1), \dots, L(u_n))$.

Começamos por calcular a primeira derivada de y_1 . Encontramos

$$(6.22) \quad y'_1 = (v, u') + (v', u).$$

Temos n funções v_1, \dots, v_n a determinar, pelo que deverá ser possível estabelecer n condições que as relacionem. Se impomos a condição de que o segundo termo no segundo membro de (6.22) deva ser nulo, a fórmula para y'_1 simplifica-se e fica

$$y'_1 = (v, u'), \quad \text{com tanto que } (v', u) = 0.$$

Derivando a relação para y'_1 obtemos

$$y''_1 = (v, u'') + (v', u').$$

Se pudermos escolher v de maneira que $(v', u') = 0$ então a fórmula para y_1' também se simplifica e escreve-se

$$y_1'' = (v, u''), \quad \text{desde que também } (v', u') = 0.$$

Se continuamos este raciocínio para as $n - 1$ primeiras derivadas de y_1 , obtemos

$$y_1^{(n-1)} = (v, u^{(n-1)}), \quad \text{desde que } (v', u^{(n-2)}) = 0$$

Obtivemos até ao momento $n - 1$ condições relativas a τ . Derivando ainda uma vez obtemos

$$y_1^{(n)} = (v, u^{(n)}) + (v', u^{(n-1)}).$$

Impondo a condição $(v', u^{(n-1)}) = R(x)$, a última equação vem

$$y_1^{(n)} = (v, u^{(n)}) + R(x), \quad \text{desde que } (v', u^{(n-1)}) = R(x).$$

Admitamos, por agora, que se podem satisfazer as n condições impostas a v . Seja $L = D^n + P_1(x)D^{n-1} + \dots + P_n(x)$. Quando aplicamos L a y_1 encontramos

$$\begin{aligned} L(y_1) &= y_1^{(n)} + P_1(x)y_1^{(n-1)} + \dots + P_n(x)y_1 \\ &= \{(v, u^{(n)}) + R(x)\} + P_1(x)(v, u^{(n-1)}) + \dots + P_n(x)(v, u) \\ &= (v, L(u)) + R(x) = (v, 0) + R(x) = R(x). \end{aligned}$$

Assim $L(y_1) = R(x)$, pelo que y_1 é uma solução da equação não homogénea.

O método terá êxito se for possível satisfazer as n condições impostas a v . Estas condições estabelecem que $(v', u^{(k)}) = 0$ para $k = 0, 1, \dots, n - 2$ e que $(v', u^{(n-1)}) = R(x)$. Podemos pois escrever estas n equações como uma simples equação matricial.

$$(6.23) \quad W(x)v'(x) = R(x) \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix},$$

onde $v'(x)$ é uma matriz coluna $n \times 1$, e W é a matriz $n \times n$ cujas linhas são formadas pelos componentes de u e suas sucessivas derivadas:

$$W = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \\ u_1' & u_2' & \cdots & u_n' \\ \vdots & \vdots & \ddots & \vdots \\ u_1^{(n-1)} & u_2^{(n-1)} & \cdots & u_n^{(n-1)} \end{bmatrix}.$$

A matriz W chama-se a *matriz wronskiana* de u_1, \dots, u_n , segundo J. M. H. Wronski (1778-1853).

Demonstraremos, na secção seguinte, que a matriz wronskiana é não singular. Portanto podemos multiplicar ambos os membros de (6.23) por $W(x)^{-1}$ para se obter

$$v'(x) = R(x)W(x)^{-1} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

Escolhamos dois pontos c e x no intervalo J considerado e integremos esta equação vectorial no intervalo de c a x ,

$$v(x) = v(c) + \int_c^x R(t)W(t)^{-1} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} dt = v(c) + z(x),$$

com

$$z(x) = \int_c^x R(t)W(t)^{-1} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} dt.$$

A fórmula $y_1 = (u, v)$ para a solução particular toma agora o aspecto

$$y_1 = (u, v) = (u, v(c) + z) = (u, v(c)) + (u, z).$$

O primeiro termo $(u, v(c))$ satisfaz à equação homogénea, visto que é uma combinação de u_1, \dots, u_n . Por conseguinte podemos omitir este termo e utilizar o segundo termo (u, z) como uma solução particular da equação não homogénea. Por outras palavras, uma solução particular de $L(y) = R$ é dada pelo produto interno

$$(u(x), z(x)) = \left(u(x), \int_c^x R(t)W(t)^{-1} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} dt \right).$$

Observe-se que não é necessário que a função R seja contínua no intervalo J . Tudo o que se exige é que R seja integrável em $[c, x]$.

Podemos resumir os resultados desta secção no seguinte teorema.

TEOREMA 6.11. *Sejam u_1, \dots, u_n n soluções independentes da equação diferencial linear homogénea de ordem n , $L(y) = 0$, num intervalo J . Uma solução particular y_1 da equação não homogénea $L(y) = R$ é dada pela fórmula*

$$y_1(x) = \sum_{k=1}^n u_k(x)v_k(x),$$

com v_1, \dots, v_n os elementos da matriz coluna v , $n \times 1$, definida pela equação

$$(6.24) \quad v(x) = \int_c^x R(t)W(t)^{-1} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} dt.$$

Nesta fórmula, W é a matriz wronskiana de u_1, \dots, u_n e c um ponto qualquer de J .

Nota: O integral definido (6.24) pode substituir-se por um integral indefinido

$$\int R(x)W(x)^{-1} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} dx.$$

EXEMPLO 1. Determinar a solução geral da equação diferencial

$$y'' - y = \frac{2}{1 + e^x}$$

no intervalo $(-\infty, +\infty)$.

Resolução. A equação homogénea $(D^2 - 1)y = 0$ tem as duas soluções independentes $u_1(x) = e^x$, $u_2(x) = e^{-x}$. A matriz wronskiana de u_1 e u_2 é

$$W(x) = \begin{bmatrix} e^x & e^{-x} \\ e^x & -e^{-x} \end{bmatrix}.$$

Porque $\det W(x) = -2$, a matriz é não singular e a sua inversa é dada por

$$W(x)^{-1} = -\frac{1}{2} \begin{bmatrix} -e^{-x} & -e^{-x} \\ -e^x & e^x \end{bmatrix}.$$

Portanto

$$W(x)^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = -\frac{1}{2} \begin{bmatrix} -e^{-x} \\ e^x \end{bmatrix}$$

e temos

$$R(x)W(x)^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = -\frac{1}{2} \frac{2}{1+e^x} \begin{bmatrix} -e^{-x} \\ e^x \end{bmatrix} = \begin{bmatrix} \frac{e^{-x}}{1+e^x} \\ \frac{-e^x}{1+e^x} \end{bmatrix}.$$

Integrando cada componente do vector no segundo membro encontramos

$$v_1(x) = \int \frac{e^{-x}}{1+e^x} dx = \int \left(e^{-x} - 1 + \frac{e^x}{1+e^x} \right) dx = -e^{-x} - x + \log(1+e^x)$$

e

$$v_2(x) = \int \frac{-e^x}{1+e^x} dx = -\log(1+e^x).$$

Consequentemente a solução geral da equação diferencial é

$$\begin{aligned} y &= c_1 u_1(x) + c_2 u_2(x) + v_1(x) u_1(x) + v_2(x) u_2(x) \\ &= c_1 e^x + c_2 e^{-x} - 1 - x e^x + (e^x - e^{-x}) \log(1+e^x). \end{aligned}$$

6.12. Não singularidade da matriz wronskiana de n soluções independentes de uma equação linear homogénea

Nesta secção vamos demonstrar que a matriz wronskiana W de n soluções independentes u_1, \dots, u_n de uma equação homogénea $L(y) = 0$ é não singular. Fazêmo-lo demonstrando que a determinante de W é uma função exponencial, a qual nunca se anula no intervalo J considerado.

Seja $w(x) = \det W(x)$ para todo x em J e suponhamos que a equação diferencial, a que satisfazem u_1, \dots, u_n , tem a forma

$$(6.25) \quad y^{(n)} + P_1(x)y^{(n-1)} + \dots + P_n(x)y = 0.$$

Então tem-se:

TEOREMA 6.12. *O determinante wronskiano satisfaz à equação diferencial de primeira ordem*

$$(6.26) \quad w' + P_1(x)w = 0$$

em J . Consequentemente se $c \in J$ tem-se

$$(6.27) \quad w(x) = w(c) \exp \left[-\int_c^x P_1(t) dt \right] \quad (\text{fórmula de Abel}). \text{ Além}$$

disso, $w(x) \neq 0$ para todo x em J .

Demonstração. Seja u o vector linha $u = (u_1, \dots, u_n)$. Uma vez que cada componente de u satisfaz à equação diferencial (6.25), o mesmo se verifica com u . As linhas da matriz wronskiana W são os vectores $u, u', \dots, u^{(n-1)}$. Consequentemente podemos escrever

$$w = \det W = \det (u, u', \dots, u^{(n-1)}).$$

A derivada de w é o determinante da matriz obtida por derivação da última linha de W (ver Exercício 8 da Secção 3.17). Quer dizer

$$w' = \det (u, u', \dots, u^{(n-2)}, u^{(n)}, u^{(n)}).$$

Multiplicando a última linha de w por $P_1(x)$ temos também

$$P_1(x)w = \det (u, u', \dots, u^{(n-2)}, P_1(x)u^{(n-1)}).$$

Somando, membro a membro, as duas últimas equações encontramos

$$w' + P_1(x)w = \det (u, u', \dots, u^{(n-2)}, u^{(n)} + P_1(x)u^{(n-1)}).$$

Mas as linhas deste último determinante são dependentes, visto que u verifica a equação diferencial (6.25). Portanto o determinante é zero, o que significa que w satisfaz a (6.26). Resolvendo (6.26) obtemos a fórmula de Abel (6.27).

Seguidamente vamos provar que $w(c) \approx 0$ para algum c em J . Fazêmo-lo por redução ao absurdo. Suponhamos que $w(t) = 0$ para todo t em J . Escolhamos um valor fixo de t em J , por exemplo $t = t_0$, e consideremos o sistema linear de equações algébricas

$$W(t_0)X = 0,$$

onde X é um vector coluna. Uma vez que $\det W(t_0) = 0$, a matriz $W(t_0)$ é singular pelo que este sistema admite uma solução não nula, seja $X = (c_1, \dots, c_n) \neq (0, 0, \dots, 0)$. Utilizando as componentes deste vector não nulo, seja f a combinação linear

$$f(t) = c_1 u_1(t) + \cdots + c_n u_n(t).$$

A função f assim definida satisfaz $L(f) = 0$ em J , visto que é uma combinação linear de u_1, \dots, u_n . A equação matricial $W(t_0)X = O$ implica que

$$f(t_0) = f'(t_0) = \cdots = f^{(n-1)}(t_0) = 0.$$

Por conseguinte f tem como valor inicial o vector O em $t = t_0$ de modo que, pelo teorema de unicidade, f é a solução zero. Significa isto que $c_1 = \cdots = c_n = 0$, o que é absurdo. Deste modo $w(t) \neq 0$, para algum t em J . Tomando para este valor de t o c na fórmula de Abel, vemos que $w(x) \neq 0$ para todo x em J . Está assim completamente demonstrado o teorema 6.12.

6.13. Métodos especiais para determinação de soluções particulares de equações não homogêneas. Redução a um sistema de equações lineares de primeira ordem

Embora o método de variação das constantes seja um método geral para determinação de uma solução particular de $L(y) = R$, é possível estabelecer métodos especiais que são muitas vezes mais fáceis de aplicar quando a equação assume formas particulares. Por exemplo, se a equação tem coeficientes constantes podemos reduzir o problema ao da resolução de uma sucessão de equações diferenciais lineares de primeira ordem. O método geral ilustra-se melhor com um exemplo simples.

EXEMPLO 1. Determinar uma solução particular da equação

$$(6.28) \quad (D - 1)(D - 2)y = xe^{x+x^2}.$$

Resolução. Seja $u = (D - 2)y$. Então a equação vem

$$(D - 1)u = xe^{x+x^2}.$$

Esta é uma equação diferencial linear de primeira ordem em u que pode ser resolvida segundo o teorema 6.1. Uma solução particular é

$$u = \frac{1}{2}e^{x+x^2}.$$

Substituindo na equação $u = (D - 2)y$ obtemos

$$(D - 2)y = \frac{1}{2}e^{x+x^2},$$

uma equação diferencial linear de primeira ordem para y . Resolvendo esta pelo teorema 6.1 verificamos que uma solução particular (com $y_1(0) = 0$) é dada por

$$y_1(x) = \frac{1}{2}e^{2x} \int_0^x e^{t^2-t} dt.$$

Muito embora o integral não possa ser expresso por meio de funções elementares, consideramos a equação diferencial como resolvida, uma vez que a solução é expressa por meio de integrais de funções elementares. A solução geral (6.28) é

$$y = c_1 e^x + c_2 e^{2x} + \frac{1}{2} e^{2x} \int_0^x e^{t^2-t} dt.$$

6.14. O método do anulador para determinação de uma solução particular da equação não homogênea

O método que vamos passar a tratar pode utilizar-se caso a equação $L(y) = R$ tenha coeficientes constantes e o segundo membro R seja anulado por um operador de coeficientes constantes, por exemplo $A(R) = 0$. Em princípio o método é muito simples. Aplicamos o operador A a ambos os membros da equação diferencial $L(y) = R$ e obtemos uma nova equação $AL(y) = 0$ a qual deve ser satisfeita por todas as soluções da equação original. Visto que AL é outro operador com coeficientes constantes, podemos determinar o seu espaço nulo pelo cálculo das raízes da equação característica de AL . Então o problema reduz-se à escolha, a partir desse espaço nulo, de uma função particular y_1 que satisfaz a $L(y_1) = R$. O exemplo seguinte esclarece a aplicação do método.

EXEMPLO 1. Determinar uma solução particular da equação

$$(D^4 - 16)y = x^4 + x + 1.$$

Resolução. O segundo membro, um polinômio do quarto grau, é anulado pelo operador D^5 . Portanto qualquer solução da equação dada é também uma solução da equação

$$(6.29) \quad D^5(D^4 - 16)y = 0.$$

As raízes da equação característica são 0, 0, 0, 0, 0, 2, -2, $2i$, $-2i$, pelo que todas as soluções de (6.29) se encontram na combinação linear

$$y = c_1 + c_2 x + c_3 x^2 + c_4 x^3 + c_5 x^4 + c_6 e^{2x} + c_7 e^{-2x} + c_8 \cos 2x + c_9 \sin 2x.$$

Desejamos escolher os c_i de maneira que $L(y) = x^4 + x + 1$, com $L = D^4 - 16$. Visto que os últimos quatro termos são anulados por L , podemos tomar $c_6 = c_7 = c_8 = c_9 = 0$ e tentar determinar c_1, \dots, c_5 de tal maneira que

$$L(c_1 + c_2 x + c_3 x^2 + c_4 x^3 + c_5 x^4) = x^4 + x + 1.$$

Por outras palavras, obtemos uma solução particular y_1 que é um polinômio do 4.º grau satisfazendo a $L(y_1) = x^4 + x + 1$. Escrevamos, para simplificar,

$$16y_1 = ax^4 + bx^3 + cx^2 + dx + e.$$

Daqui concluímos $16y_1^{(4)} = 24a$, donde $y_1^{(4)} = 3a/2$. Substituindo na equação diferencial $L(y_1) = x^4 + x + 1$, temos que determinar a, b, c, d, e que satisfaçam a

$$\frac{3}{2}a - ax^4 - bx^3 - cx^2 - dx - e = x^4 + x + 1.$$

Igualando os coeficientes dos termos semelhantes obtemos

$$a = -1, \quad b = c = 0, \quad d = -1, \quad e = -\frac{5}{2},$$

pelo que a solução particular y_1 é dada por

$$y_1 = -\frac{1}{16}x^4 - \frac{1}{16}x - \frac{5}{32}.$$

EXEMPLO 2. Resolver a equação diferencial $y'' - 5y' + 6y = xe^x$.

Resolução. A equação diferencial é da forma

$$(6.30) \quad L(y) = R,$$

onde $R(x) = xe^x$ e $L = D^2 - 5D + 6$. A correspondente equação homogénea escreve-se

$$(D - 2)(D - 3)y = 0;$$

e admite as soluções independentes $u_1(x) = e^{2x}$, $u_2(x) = e^{3x}$. Interessa-nos agora uma solução particular y_1 da equação não homogénea. Verificamos que a função $R(x) = xe^x$ é uma solução da equação homogénea

$$(D - 1)^2 y = 0.$$

Portanto, se aplicamos a ambos os membros de (6.30) o operador $(D - 1)^2$, encontramos que qualquer função que satisfaça (6.30) deve também satisfazer à equação

$$(D - 1)^2(D - 2)(D - 3)y = 0.$$

Esta equação diferencial tem as raízes características 1, 1, 2, 3, motivo porque todas as suas soluções se encontram na combinação linear

$$y = ae^x + bxe^x + ce^{2x} + de^{3x},$$

onde a, b, c, d são constantes. Pretendemos escolher a, b, c, d de tal maneira que a solução y_1 verifique $L(y_1) = xe^x$. Visto que $L(ce^{2x} + de^{3x}) = 0$, quaisquer que sejam

os c e d necessitamos unicamente escolher a e b pelo que $L(ae^x + bxe^x) = xe^x$ e tomamos $c = d = 0$. Se fizermos

$$y_1 = ae^x + bxe^x,$$

temos

$$D(y_1) = (a + b)e^x + bxe^x, \quad D^2(y_1) = (a + 2b)e^x + bxe^x,$$

pelo que a equação $(D^2 - 5D + 6)y_1 = xe^x$ vem

$$(2a - 3b)e^x + 2bxe^x = xe^x.$$

Dividindo por e^x e igualando os coeficientes dos termos semelhantes encontramos $a = \frac{3}{4}$, $b = \frac{1}{4}$. Assim $y_1 = \frac{3}{4}e^x + \frac{1}{4}xe^x$ e a solução geral de $L(y) = R$ é dada pela fórmula

$$y = c_1e^{2x} + c_2e^{3x} + \frac{3}{4}e^x + \frac{1}{4}xe^x.$$

O método usado nos exemplos anteriores chama-se o *método do anulador*. Será sempre aplicável se podermos determinar um operador de coeficientes constantes A que anule R . Do nosso conhecimento das equações diferenciais lineares homogêneas com coeficientes constantes sabemos que as únicas funções reais anuladas por operadores com coeficientes constantes são combinações lineares de funções da forma

$$x^{m-1}e^{\alpha x}, \quad x^{m-1}e^{\alpha x} \cos \beta x, \quad x^{m-1}e^{\alpha x} \sin \beta x,$$

onde m é um inteiro positivo e α e β são constantes reais. A função $y = x^{m-1}e^{\alpha x}$ é uma solução da equação diferencial com uma raiz característica α com multiplicidade m . Portanto, esta função tem o anulador $(D - \alpha)^m$. Cada uma das funções $y = x^{m-1}e^{\alpha x} \cos \beta x$ e $y = x^{m-1}e^{\alpha x} \sin \beta x$ é uma solução de uma equação diferencial com raízes características completas $\alpha \pm i\beta$, cada uma ocorrendo com multiplicidade m , pelo que são anuladas pelo operador $[D^2 - 2\alpha D + (\alpha^2 + \beta^2)]^m$. Para facilitar a resolução de alguns exercícios, apresentamos a seguir uma lista desses anuladores na tabela 6.1, junto com alguns dos seus casos particulares.

TABELA 6.1.

Função	Anulador
$y = x^{m-1}$	D^m
$y = e^{\alpha x}$	$D - \alpha$
$y = x^{m-1}e^{\alpha x}$	$(D - \alpha)^m$
$y = \cos \beta x$ ou $y = \sin \beta x$	$D^2 + \beta^2$
$y = x^{m-1} \cos \beta x$ ou $y = x^{m-1} \sin \beta x$	$(D^2 + \beta^2)^m$
$y = e^{\alpha x} \cos \beta x$ ou $y = e^{\alpha x} \sin \beta x$	$D^2 - 2\alpha D + (\alpha^2 + \beta^2)$
$y = x^{m-1}e^{\alpha x} \cos \beta x$ ou $y = x^{m-1}e^{\alpha x} \sin \beta x$	$[D^2 - 2\alpha D + (\alpha^2 + \beta^2)]^m$

Embora o método do anulador seja muito eficiente quando aplicável, está limitada a sua utilização a equações cujos segundos membros R tenham um anulador com coeficientes constantes. Se $R(x)$ tem a forma e^{ax} , $\log x$, ou $\operatorname{tg} x$, o método não é utilizável; devemos neste caso recorrer ao método de variação das constantes, ou algum outro método para determinar uma solução particular.

6.15. Exercícios

Em cada um dos Exercícios 1 a 10, determinar a solução geral no intervalo $(-\infty, +\infty)$.

1. $y'' - y' = x^2$.
2. $y'' - 4y = e^{2x}$.
3. $y'' + 2y' = 3xe^x$.
4. $y'' + 4y = \sin x$.
5. $y'' - 2y' + y = e^x + e^{2x}$.
6. $y'' - y' = e^x$.
7. $y'' - y' = e^x + e^{-x}$.
8. $y'' + 3y' + 3y' + y = xe^{-x}$.
9. $y'' + y = xe^x \sin 2x$.
10. $y^{(4)} - y = x^2 e^{-x}$.

11. Se um operador A , de coeficientes constantes, anula f e outro operador de coeficientes constantes, B , anula g , mostrar que o produto AB anula $f + g$.
12. Seja A um operador de coeficientes constantes com polinômio característico p_A .
(a) Utilizar o método do anulador para provar que a equação diferencial $A(y) = e^{ax}$ tem uma solução particular da forma

$$y_1 = \frac{e^{ax}}{p_A(\alpha)}$$

se α não for raiz do polinômio p_A .

- (b) Se α é uma raiz simples de p_A (multiplicidade 1), provar que a equação $A(y) = e^{ax}$ tem a solução particular

$$y_1 = \frac{xe^{ax}}{p'_A(\alpha)}$$

- (c) Generalizar os resultados de (a) e (b) quando α é um zero de p_A com multiplicidade m .
13. São dados dois operadores de coeficientes constantes A e B cujos polinômios característicos não tem zeros em comum. Seja $C = AB$.
(a) Provar que toda a solução da equação diferencial $C(y) = 0$ tem a forma $y = y_1 + y_2$, onde $A(y_1) = 0$ e $B(y_2) = 0$.
(b) Demonstrar que as funções y_1 e y_2 na alínea (a) são univocamente determinadas, isto é, para um dado y verificando a condição $C(y) = 0$ existe somente um par y_1, y_2 com as propriedades da alínea (a).
14. Se $L(y) = y'' + ay' + by$, com a e b constantes, seja f uma solução particular de $L(y) = 0$ satisfazendo às condições $f(0) = 0$ e $f'(0) = 1$. Mostrar que uma solução particular de $L(y) = R$ é dada pela fórmula

$$y_1(x) = \int_0^x f(x-t)R(t) dt$$

para qualquer escolha de c . Em particular, se as raízes da equação característica são iguais, por exemplo $r_1 = r_2 = m$, mostrar que a fórmula para $y_1(x)$ vem

$$y_1(x) = e^{mx} \int_0^x (x-t)e^{-mt}R(t) dt.$$

15. Seja Q o operador "multiplicação por x ", isto é, $Q(y)(x) = x \cdot y(x)$, para todo y da classe \mathcal{C}^∞ e todo real x . Designe-se por I o operador identidade, definido por $I(y) = y$ para todo y em \mathcal{C}^∞ .
- (a) Provar que $DQ - QD = I$.
- (b) Mostrar que $D^2Q - QD^2$ é um operador de coeficientes constantes de primeira ordem, e determinar este operador explicitamente como um polinômio linear em D .
- (c) Mostrar que $D^3Q - QD^3$ é um operador de coeficientes constantes de segunda ordem, e determinar este operador explicitamente como um polinômio do segundo grau em D .
- (d) Enunciar a generalização sugerida pelo operador $D^nQ - QD^n$ e demonstrá-la por indução.

Em cada um dos Exercícios 16 a 20, determinar a solução geral da equação diferencial no intervalo dado.

16. $y'' - y = 1/x$, $(0, +\infty)$.
17. $y'' + 4y = \sec 2x$, $\left(-\frac{\pi}{4}, \frac{\pi}{4}\right)$.
18. $y'' - y = \sec^3 x - \sec x$, $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$.
19. $y'' - 2y' + y = e^{e^x}(e^x - 1)^2$, $(-\infty, +\infty)$.
20. $y''' - 7y'' + 14y' - 8y = \log x$, $(0, +\infty)$.

6.16. Exercícios variados sobre equações diferenciais lineares

1. Uma curva integral $y = u(x)$, da equação diferencial $y'' - 3y' - 4y = 0$, intersecta na origem uma curva integral $y = v(x)$, da equação diferencial $y'' + 4y' - 5y = 0$. Determinar as funções u e v se as duas curvas tem iguais declives na origem e se

$$\lim_{x \rightarrow \infty} \frac{[v(x)]^4}{u(x)} = \frac{5}{6}.$$

2. Uma curva integral $y = u(x)$ da equação diferencial $y'' - 4y' + 29y = 0$ intersecta uma curva integral $y = v(x)$ da equação diferencial $y'' + 4y' + 13y = 0$ na origem. As duas curvas têm iguais declives na origem. Determinar u e v se $u(\pi/2) = 1$.
3. Sabendo que a equação diferencial $y'' + 4xy' + Q(x)y = 0$ tem duas soluções da forma $y_1 = u(x)$ e $y_2 = xu(x)$, onde $u(0) = 1$, determinar $u(x)$ e $Q(x)$ explicitamente em funções de x .
4. Seja $L(y) = y'' + P_1y' + P_2y$. Para se resolver a equação não homogênea $L(y) = R$ pelo método de variação dos constantes, necessitamos conhecer duas soluções linearmente independentes da equação homogênea. Este exercício mostra que se fôr conhecida uma solução u_1 de $L(y) = 0$, e se u_1 nunca se anula no intervalo J , uma segunda solução u_2 da equação homogênea é dada por

$$u_2(x) = u_1(x) \int_c^x \frac{Q(t)}{[u_1(t)]^2} dt,$$

onde $Q(x) = e^{-\int P_1(x) dx}$, e c é um ponto qualquer de J . Estas duas soluções são independentes em J .

- (a) Provar que a função u_2 satisfaz $L(y) = 0$.
 (b) Provar que u_1 e u_2 são independentes em J .
 5. Determinar a solução geral da equação

$$xy'' - 2(x+1)y' + (x+2)y = x^3e^{2x}$$

para $x > 0$, dado que a equação homogénea tem uma solução da forma $y = e^{mx}$.

6. Por análise directa da equação obter uma solução não nula e determinar depois a solução geral da equação diferencial

$$(y'' - 4y') + x^2(y' - 4y) = 0.$$

7. Determinar a solução geral da equação diferencial

$$4x^2y'' + 4xy' - y = 0,$$

dado que existe uma solução particular da forma $y = x^m$ para $x > 0$.

8. Determinar uma solução da equação homogénea por tentativas e determinar em seguida a solução geral da equação

$$x(1-x)y'' - (1-2x)y' + (x^2-3x+1)y = (1-x)^3.$$

9. Determinar a solução geral da equação

$$(2x - 3x^3)y'' + 4y' + 6xy = 0,$$

sabendo que tem uma solução que é um polinómio em x .

10. Determinar a solução geral da equação

$$x^2(1-x)y'' + 2x(2-x)y' + 2(1+x)y = x^2,$$

sabendo que a correspondente equação homogénea tem uma solução da forma $y = x^c$.

11. Seja $g(x) = \int_x^t e^{t/t} dt$ se $x > 0$. (Não tentar calcular o integral.) Determinar todos os valores da constante a tais que a função f definida por

$$f(x) = \frac{1}{x} e^{ag(x)}$$

verifica a equação diferencial linear

$$x^2y'' + (3x - x^2)y' + (1 - x - e^{2x})y = 0.$$

Uma vez sabido isto, determinar a solução geral da equação no intervalo $(0, +\infty)$.

6.17. Equações lineares de segunda ordem com coeficientes analíticos

Uma função f diz-se analítica no intervalo $(x_0 - r, x_0 + r)$ se f admite um desenvolvimento em série de potências nesse intervalo,

$$f(x) = \sum_{n=0}^{\infty} a_n(x - x_0)^n,$$

convergente para $|x - x_0| < r$. Se os coeficientes de uma equação diferencial linear homogênea

$$y^{(n)} + P_1(x)y^{(n-1)} + \cdots + P_n(x)y = 0$$

são analíticos num intervalo $(x_0 - r, x_0 + r)$, então pode demonstrar-se que existem n soluções independentes u_1, \dots, u_n , cada uma das quais é analítica no mesmo intervalo. Vamos demonstrar este teorema para equações de segunda ordem e discutiremos depois um exemplo importante que se apresenta em muitas aplicações.

TEOREMA 6.13. *Se P_1 e P_2 são analíticas num intervalo aberto $(x_0 - r, x_0 + r)$, portanto*

$$P_1(x) = \sum_{n=0}^{\infty} b_n(x - x_0)^n, \quad P_2(x) = \sum_{n=0}^{\infty} c_n(x - x_0)^n,$$

então a equação diferencial

$$(6.31) \quad y'' + P_1(x)y' + P_2(x)y = 0$$

tem duas soluções independentes u_1 e u_2 as quais são analíticas no mesmo intervalo.

Demonstração. Pretendemos encontrar uma solução na forma de série de potências

$$(6.32) \quad y = \sum_{n=0}^{\infty} a_n(x - x_0)^n,$$

convergente no intervalo dado. Para isso substituímos as séries dadas para P_1 e P_2 na equação diferencial e determinamos em seguida relações às quais devem satisfazer os coeficientes a_n de modo a que a função y , dada por (6.32), satisfaça a equação.

As derivadas y' e y'' podem obter-se derivando, termo a termo, o desenvolvimento de y em série de potências (ver teorema 11.9, Volume I). Obtemos pois

$$y' = \sum_{n=1}^{\infty} n a_n(x - x_0)^{n-1} = \sum_{n=0}^{\infty} (n+1) a_{n+1}(x - x_0)^n,$$

$$y'' = \sum_{n=2}^{\infty} n(n-1) a_n(x - x_0)^{n-2} = \sum_{n=0}^{\infty} (n+2)(n+1) a_{n+2}(x - x_0)^n.$$

Os produtos $P_1(x)y'$ e $P_2(x)y$ são dados pelas séries de potências†

$$P_1(x)y' = \sum_{n=0}^{\infty} \left(\sum_{k=0}^n (k+1)a_{k+1}b_{n-k} \right) (x-x_0)^n$$

e

$$P_2(x)y = \sum_{n=0}^{\infty} \left(\sum_{k=0}^n a_k c_{n-k} \right) (x-x_0)^n.$$

Quando estas séries se substituem na equação diferencial (6.31) encontramos

$$\sum_{n=0}^{\infty} \left\{ (n+2)(n+1)a_{n+2} + \sum_{k=0}^n [(k+1)a_{k+1}b_{n-k} + a_k c_{n-k}] \right\} (x-x_0)^n = 0.$$

Portanto a equação diferencial verificar-se-á se escolhemos os coeficientes a_n de maneira que a fórmula

$$(6.33) \quad (n+2)(n+1)a_{n+2} = - \sum_{k=0}^n [(k+1)a_{k+1}b_{n-k} + a_k c_{n-k}]$$

para $n = 0, 1, 2, \dots$. Esta fórmula exprime a_{n+2} por intermédio dos coeficientes procedentes a_0, a_1, \dots, a_{n+1} e os coeficientes das funções P_1 e P_2 . Escolhamos valores arbitrários para os dois primeiros coeficientes a_0 e a_1 e utilizemos (6.33) para determinar os restantes coeficientes a_2, a_3, \dots , em função de a_0 e a_1 . Este processo garante que a série de potências (6.32) satisfará à equação diferencial (6.31). A fase imediata na demonstração consiste em provar que a série assim definida converge para todo x no intervalo $(x_0 - r, x_0 + r)$. Isto pode fazer-se majorando a série (6.32) por outra série de potências que se saiba ser convergente. Finalmente, demonstramos que podemos escolher a_0 e a_1 para obter duas soluções independentes.

Provemos então que a série (6.32) cujos coeficientes são definidos por (6.33) converge no intervalo dado.

Seja $x_1 \neq x_0$ um ponto fixo no intervalo $(x_0 - r, x_0 + r)$ e $t = |x_1 - x_0|$. Uma vez que as séries para P_1 e P_2 convergem absolutamente para $x = x_1$, os termos destas séries são limitados, isto é

$$|b_k| t^k \leq M_1 \quad \text{e} \quad |c_k| t^k \leq M_2,$$

para algum $M_1 > 0$ e algum $M_2 > 0$. Seja M o maior dos valores M_1 e tM_2 . Então temos

$$|b_k| \leq \frac{M}{t^k} \quad \text{e} \quad |c_k| \leq \frac{M}{t^{k+1}}.$$

A fórmula de recorrência implica a desigualdade

† Os leitores não familiarizados com o produto de séries de potências podem consultar o Exercício 7 da Secção 6.21.

$$\begin{aligned}
 (n+2)(n+1)|a_{n+2}| &\leq \sum_{k=0}^n \left\{ (k+1)|a_{k+1}| \frac{M}{t^{n-k}} + |a_k| \frac{M}{t^{n-k+1}} \right\} \\
 &= \frac{M}{t^{n+1}} \left\{ \sum_{k=0}^n (k+1)|a_{k+1}| t^{k+1} + \sum_{k=0}^n |a_{k+1}| t^{k+1} + |a_0| - |a_{n+1}| t^{n+1} \right\} \\
 &\leq \frac{M}{t^{n+1}} \left\{ \sum_{k=0}^n (k+2)|a_{k+1}| t^{k+1} + |a_0| \right\} = \frac{M}{t^{n+1}} \sum_{k=0}^{n+1} (k+1)|a_k| t^k.
 \end{aligned}$$

Façamos agora $A_0 = |a_0|$, $A_1 = |a_1|$, e definamos A_2, A_3, \dots sucessivamente pela fórmula de recorrência

$$(6.34) \quad (n+2)(n+1)A_{n+2} = \frac{M}{t^{n+1}} \sum_{k=0}^{n+1} (k+1)A_k t^k$$

para $n \geq 0$. Então $|a_n| \leq A_n$ para todo $n \geq 0$, pelo que a série $\sum a_n(x-x_0)^n$ é dominada pela série $\sum A_n |x-x_0|^n$. Aplicamos agora o critério do cociente para provar que $\sum A_n |x-x_0|^n$ converge se $|x-x_0| < t$.

Substituindo n por $n-1$ em (6.34) e subtraindo t^{-1} vezes a equação resultante de (6.34) encontramos que $(n+2)(n+1)A_{n+2} - t^{-1}(n+1)nA_{n+1} = M(n+2)A_{n+1}$. Portanto

$$A_{n+2} = A_{n+1} \frac{(n+1)n + (n+2)Mt}{(n+2)(n+1)t},$$

e encontramos

$$\frac{A_{n+2}|x-x_0|^{n+2}}{A_{n+1}|x-x_0|^{n+1}} = \frac{(n+1)n + (n+2)Mt}{(n+2)(n+1)t} |x-x_0| \rightarrow \frac{|x-x_0|}{t}$$

quando $n \rightarrow \infty$. Este limite é menor que 1 se $|x-x_0| < t$. Por isso $\sum a_n(x-x_0)^n$ converge se $|x-x_0| < t$. Mas porque $t = |x_1 - x_0|$ é uma vez que x_1 era um ponto arbitrário no intervalo $(x_0 - r, x_0 + r)$, a série $\sum a_n(x-x_0)^n$ converge para todo x em $(x_0 - r, x_0 + r)$.

Os dois primeiros coeficientes a_0 e a_1 representam os valores iniciais de y e da sua derivada no ponto x_0 . Se u_1 é a série de potências solução com $a_0 = 1$ e $a_1 = 0$, de maneira que

$$u_1(x_0) = 1 \quad \text{e} \quad u_1'(x_0) = 0,$$

e u_2 a solução com $a_0 = 0$ e $a_1 = 1$, de maneira que

$$u_2(x_0) = 0 \quad \text{e} \quad u_2'(x_0) = 1,$$

então as soluções u_1 e u_2 serão independentes e o teorema está demonstrado.

6.18. A equação de Legendre

Nesta secção vamos encontrar soluções, em forma de séries de potências, para a equação de Legendre,

$$(6.35) \quad (1 - x^2)y'' - 2xy' + \alpha(\alpha + 1)y = 0,$$

com α qualquer constante real. Esta equação aparece em problemas de atracção e de fluxo de calor com simetria esférica. Quando α é um inteiro positivo encontraremos para a equação soluções sob a forma de polinómios chamados *polinómios de Legendre*. Estes são os mesmos polinómios que encontramos atrás, em ligação com o método de Gram-Schmidt (Capítulo 1, pg. 29).

A equação de Legendre pode escrever-se

$$[(x^2 - 1)y']' = \alpha(\alpha + 1)y,$$

o que significa ser da forma

$$T(y) = \lambda y,$$

sendo T um operador de Sturm-Liouville, $T(fy) = (pf)'$, com $p(x) = x^2 - 1$ e $\lambda = \alpha(\alpha + 1)$.

Por conseguinte, as soluções não nulas da equação de Legendre são funções próprias de T correspondentes ao valor próprio $\alpha(\alpha + 1)$. Posto que $p(x)$ satisfaz às condições de fronteira.

$$p(1) = p(-1) = 0,$$

o operador T é simétrico relativamente ao produto interno

$$(f, g) = \int_{-1}^1 f(x)g(x) dx.$$

A teoria geral dos operadores simétricos diz-nos que as funções próprias correspondentes a valores próprios distintos são ortogonais (teorema 5.3).

Na equação diferencial tratada no teorema 6.13 o coeficiente de y'' é 1. A equação de Legendre pode escrever-se nesta forma se dividirmos ambos os membros por $1 - x^2$. De (6.35) resulta

$$y'' + P_1(x)y' + P_2(x)y = 0,$$

com

$$P_1(x) = -\frac{2x}{1 - x^2} \quad \text{e} \quad P_2(x) = \frac{\alpha(\alpha + 1)}{1 - x^2},$$

se $x^2 \neq 1$. Uma vez que $1/(1 - x^2) = \sum_{n=0}^{\infty} x^{2n}$ para $|x| < 1$, ambos P_1 e P_2 admitem desenvolvimentos em séries de potências no intervalo $(-1, +1)$ pelo que o teorema 6.13 é aplicável. Para encontrar a fórmula por recorrência para os coeficientes, é mais simples deixar a equação na forma (6.35) e tentar encontrar uma série de potências solução da forma

$$y = \sum_{n=0}^{\infty} a_n x^n$$

válida no intervalo aberto $(-1, 1)$. Derivando esta série, termo a termo, obtemos

$$y' = \sum_{n=1}^{\infty} n a_n x^{n-1} \quad \text{e} \quad y'' = \sum_{n=2}^{\infty} n(n-1) a_n x^{n-2}.$$

Deste modo temos

$$2xy' = \sum_{n=1}^{\infty} 2n a_n x^n = \sum_{n=0}^{\infty} 2n a_n x^n,$$

e

$$\begin{aligned} (1-x^2)y'' &= \sum_{n=2}^{\infty} n(n-1) a_n x^{n-2} - \sum_{n=2}^{\infty} n(n-1) a_n x^n \\ &= \sum_{n=0}^{\infty} (n+2)(n+1) a_{n+2} x^n - \sum_{n=0}^{\infty} n(n-1) a_n x^n \\ &= \sum_{n=0}^{\infty} [(n+2)(n+1) a_{n+2} - n(n-1) a_n] x^n. \end{aligned}$$

Se substituirmos estas séries na equação diferencial (6.35), concluímos que a equação verificar-se-á se, e só se, os coeficientes verificarem a relação

$$(n+2)(n+1) a_{n+2} - n(n-1) a_n - 2n a_n + \alpha(\alpha+1) a_n = 0$$

para todo $n \geq 0$. Esta equação é a mesma que

$$(n+2)(n+1) a_{n+2} - (n-\alpha)(n+1+\alpha) a_n = 0,$$

ou

$$(6.36) \quad a_{n+2} = - \frac{(\alpha-n)(\alpha+n+1)}{(n+1)(n+2)} a_n.$$

Esta relação permite-nos determinar a_2, a_4, a_6, \dots , sucessivamente em função de a_0 . Analogamente, podemos calcular a_3, a_5, a_7, \dots , em função de a_1 . Para os coeficientes com índices pares temos

$$a_2 = - \frac{\alpha(\alpha+1)}{1 \cdot 2} a_0,$$

$$a_4 = - \frac{(\alpha-2)(\alpha+3)}{3 \cdot 4} a_2 = (-1)^2 \frac{\alpha(\alpha-2)(\alpha+1)(\alpha+3)}{4!} a_0,$$

e, em geral,

$$a_{2n} = (-1)^n \frac{\alpha(\alpha-2) \cdots (\alpha-2n+2) \cdot (\alpha+1)(\alpha+3) \cdots (\alpha+2n-1)}{(2n)!} a_0.$$

podendo esta fórmula demonstrar-se por indução. Para os coeficientes com índices ímpares encontramos

$$a_{2n+1} = (-1)^n \frac{(\alpha - 1)(\alpha - 3) \cdots (\alpha - 2n + 1) \cdot (\alpha + 2)(\alpha + 4) \cdots (\alpha + 2n)}{(2n + 1)!} a_1.$$

Por conseguinte, a série que define y pode escrever-se

$$(6.37) \quad y = a_0 u_1(x) + a_1 u_2(x),$$

onde

$$(6.38) \quad u_1(x) = 1 + \sum_{n=1}^{\infty} (-1)^n \frac{\alpha(\alpha - 2) \cdots (\alpha - 2n + 2) \cdot (\alpha + 1)(\alpha + 3) \cdots (\alpha + 2n - 1)}{(2n)!} x^{2n}$$

e

$$(6.39) \quad u_2(x) = x + \sum_{n=1}^{\infty} (-1)^n \frac{(\alpha - 1)(\alpha - 3) \cdots (\alpha - 2n + 1) \cdot (\alpha + 2)(\alpha + 4) \cdots (\alpha + 2n)}{(2n + 1)!} x^{2n+1}.$$

O critério do cociente mostra que cada uma destas séries converge para $|x| < 1$. Também, visto ser a relação (6.36) satisfeita separadamente pelas coeficientes pares e ímpares, cada uma das funções u_1 e u_2 é uma solução de equações diferenciais (6.35). Estas soluções satisfazem às condições iniciais

$$u_1(0) = 1, \quad u_1'(0) = 0, \quad u_2(0) = 0, \quad u_2'(0) = 1.$$

Visto u_1 e u_2 serem independentes, a solução geral da equação de Legendre (6.35) no intervalo aberto $(-1, 1)$ é dada pela combinação linear (6.37) com a_1 e a_2 constantes arbitrárias.

Quando α é 0 ou um inteiro par positivo, seja $\alpha = 2m$, a série para $u_1(x)$ transforma-se num polinómio de grau $2m$ contendo unicamente potências pares de x . Uma vez que se tem

$$\alpha(\alpha - 2) \cdots (\alpha - 2n + 2) = 2m(2m - 2) \cdots (2m - 2n + 2) = \frac{2^n m!}{(m - n)!}$$

e

$$\begin{aligned} (\alpha + 1)(\alpha + 3) \cdots (\alpha + 2n - 1) &= (2m + 1)(2m + 3) \cdots (2m + 2n - 1) \\ &= \frac{(2m + 2n)! m!}{2^n (2m)! (m + n)!} \end{aligned}$$

a fórmula para $u_1(x)$ neste caso escreve-se

$$(6.40) \quad u_1(x) = 1 + \frac{(m!)^2}{(2m)!} \sum_{k=1}^m (-1)^k \frac{(2m + 2k)!}{(m - k)! (m + k)! (2k)!} x^{2k}.$$

Por exemplo, quando $\alpha = 0, 2, 4, 6$ ($m = 0, 1, 2, 3$) os polinómios correspondentes são

$$u_1(x) = 1, \quad 1 - 3x^3, \quad 1 - 10x^2 + \frac{35}{3}x^4, \quad 1 - 21x^2 + 63x^4 - \frac{231}{5}x^6.$$

A série para $u_2(x)$ não é um polinómio quando α é par, devido ao coeficiente de x^{2n+1} nunca se anular.

Quando α é um inteiro ímpar positivo, os papéis de u_1 e u_2 aparecem trocados: a série para $u_2(x)$ vem um polinómio e a série para $u_1(x)$ não é um polinómio. Concretamente se $\alpha = 2m + 1$ temos

$$(6.41) \quad u_2(x) = x + \frac{(m!)^2}{(2m+1)!} \sum_{k=1}^m (-1)^k \frac{(2m+2k+1)!}{(m-k)!(m+k)!(2k+1)!} x^{2k+1}.$$

Por exemplo, quando $\alpha = 1, 3, 5$ ($m = 0, 1, 2$), os correspondentes polinómios são

$$u_2(x) = x, \quad x - \frac{5}{3}x^3, \quad x - \frac{14}{3}x^3 + \frac{25}{5}x^5.$$

6.19. Os polinómios de Legendre

Algumas das propriedades das soluções polinomiais da equação de Legendre podem deduzir-se directamente da equação diferencial ou das fórmulas (6.40) e (6.41). Outras deduzem-se mais facilmente partindo de outra fórmula para esses polinómios que vamos deduzir agora.

Vamos estabelecer, em primeiro lugar, uma fórmula simples que contém (a menos de factores constantes) os polinómios (6.40) e (6.41). Seja

$$(6.42) \quad P_n(x) = \frac{1}{2^n} \sum_{r=0}^{[n/2]} \frac{(-1)^r (2n-2r)!}{r! (n-r)! (n-2r)!} x^{n-2r},$$

onde $[n/2]$ representa o maior inteiro $\leq n/2$. Vamos provar que este é o *polinómio de Legendre* de grau n introduzido no capítulo 1. Quando n é par é o produto de uma constante pelo polinómio $u_1(x)$ dado em (6.40); quando n é ímpar, é o produto de uma constante pelo polinómio $u_2(x)$ dado em (6.41)† Os sete primeiros polinómios de Legendre são definidos por

$$P_0(x) = 1, \quad P_1(x) = x, \quad P_2(x) = \frac{1}{2}(3x^2 - 1), \quad P_3(x) = \frac{1}{2}(5x^3 - 3x),$$

$$P_4(x) = \frac{1}{8}(35x^4 - 30x^2 + 3), \quad P_5(x) = \frac{1}{8}(63x^5 - 70x^3 + 15x),$$

$$P_6(x) = \frac{1}{16}(231x^6 - 315x^4 + 105x^2 - 5).$$

† Quando n é par, por exemplo $n = 2m$, podemos substituir o índice de somação k na equação (6.40) por um novo índice r , onde $r = m - k$; encontramos que a soma (6.40) é o produto de uma constante por $P_n(x)$. Analogamente, quando n é ímpar, uma mudança de índice transforma a soma (6.40) no produto de uma constante por $P_n(x)$.

A figura 6.1 representa os gráficos das cinco primeiras dessas funções no intervalo $[-1, 1]$.

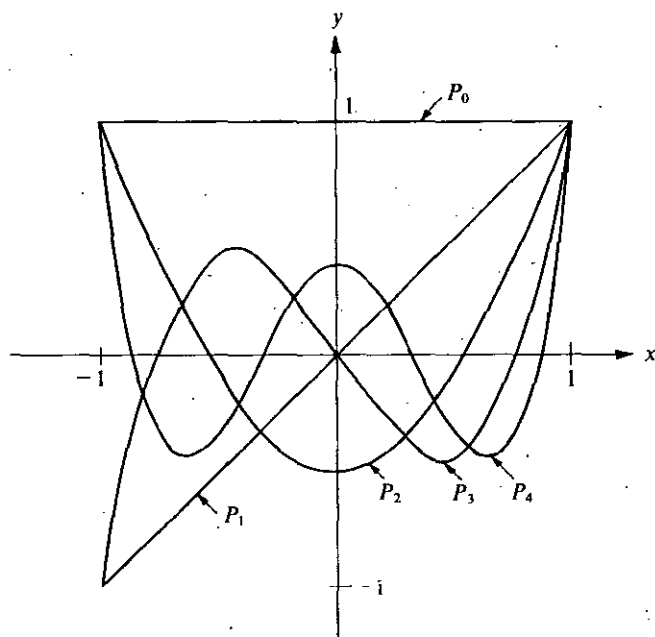


FIG. 6.1. Gráficos dos polinômios de Legendre no intervalo $[-1, 1]$.

Podemos agora demonstrar que, a menos de factores escalares, os polinômios de Legendre são os obtidos por aplicação do método de ortogonalização de Gram-Schmidt aos polinômios $1, x, x^2, \dots$, com o produto interno

$$(f, g) = \int_{-1}^1 f(x)g(x) dx.$$

Em primeiro lugar observe-se que se $m \neq n$ os polinômios P_n e P_m são ortogonais porque são funções próprias de um operador simétrico correspondendo a valores próprios distintos. Também, uma vez que P_n tem grau n e $P_0 = 1$, os polinômios $P_0(x), P_1(x), \dots, P_n(x)$ geram o mesmo subespaço que $1, x, \dots, x^n$. Na Secção 1.14, Exemplo 2, construímos outro conjunto ortogonal de polinômios y_0, y_1, y_2, \dots , tais que $y_0(x), y_1(x), \dots, y_n(x)$ geram o mesmo subespaço que $1, x, \dots, x^n$ para todo n . O teorema de ortogonalização (teorema 1.13) diz-nos que, excepto para factores escalares, existe unicamente um conjunto de funções ortogonais com esta propriedade. Por tal facto devemos ter

$$P_n(x) = c_n y_n(x)$$

para algum escalar c_n . O coeficiente de x^n em $y_n(x)$ é 1, de maneira que c_n é o coeficiente de x^n em $P_n(x)$. De (6.42) vemos que

$$c_n = \frac{(2n)!}{2^n(n!)^2}.$$

6.20. Fórmula de Rodrigues para os polinómios de Legendre

Na soma (6.42), definindo $P_n(x)$, observamos que

$$\frac{(2n-2r)!}{(n-2r)!} x^{n-2r} = \frac{d^n}{dx^n} x^{2n-2r} \quad \text{e} \quad \frac{1}{r!(n-r)!} = \frac{1}{n!} \binom{n}{r},$$

em que $\binom{n}{r}$ é o coeficiente binomial, e escrevemos a soma na forma

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} \sum_{r=0}^{\lfloor n/2 \rfloor} (-1)^r \binom{n}{r} x^{2n-2r}.$$

Quando $\lfloor n/2 \rfloor < r \leq n$, o termo x^{2n-2r} tem grau inferior a n , e portanto a sua derivada de ordem n é zero. Deste modo não se altera a soma se fizermos com que r tome os valores de 0 a n . Isto dá-nos

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} \sum_{r=0}^n (-1)^r \binom{n}{r} x^{2n-2r}.$$

Vemos agora que a soma do segundo membro é o desenvolvimento do binómio $(x^2 - 1)^n$. Quer dizer que podemos escrever

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} (x^2 - 1)^n.$$

Esta é a conhecida *fórmula de Rodrigues*, em homenagem a Olinde Rodrigues (1794-1851), um economista e reformador francês.

Com a fórmula de Rodrigues e a equação diferencial, podemos derivar um número importante de propriedades dos polinómios de Legendre. Algumas dessas propriedades apresentam-se a seguir. As correspondentes demonstrações serão efectuadas no próximo conjunto de exercícios.

Para cada $n \geq 0$ temos

$$P_n(1) = 1.$$

Além disso, $P_n(x)$ é o único polinómio que satisfaz à equação de Legendre

$$(1 - x^2)y'' - 2xy' + n(n+1)y = 0$$

e tem o valor 1 quando $x = 1$.

Para cada $n \geq 0$ temos

$$P_n(-x) = (-1)^n P_n(x).$$

Isto prova que P_n é uma função par quando n é par, e uma função ímpar quando n é ímpar.

Já mencionámos a relação de ortogonalidade

$$\int_{-1}^1 P_n(x) P_m(x) dx = 0 \quad \text{se } m \neq n.$$

Quando $m = n$ temos a expressão da norma

$$\|P_n\|^2 = \int_{-1}^1 [P_n(x)]^2 dx = \frac{2}{2n+1}.$$

Todo o polinómio de grau n pode ser expresso como uma combinação linear de polinómios de Legendre P_0, P_1, \dots, P_n . Com efeito, se f é um polinómio de grau n temos

$$f(x) = \sum_{k=0}^n c_k P_k(x),$$

onde

$$c_k = \frac{2k+1}{2} \int_{-1}^1 f(x) P_k(x) dx.$$

A partir da relação de ortogonalidade resulta que

$$\int_{-1}^1 g(x) P_n(x) dx = 0$$

para todo o polinómio g de grau menor que n . Esta propriedade pode usar-se para provar que os polinómios de Legendre P_n têm n zeros reais e distintos situados no intervalo $(-1, 1)$.

6.21. Exercícios

1. A equação de Legendre (6.35) com $\alpha = 0$ tem uma solução em forma de polinómio $u_1(x) = 1$ e a solução u_2 , não um polinómio dada pela série na equação (6.41).
(a) Mostrar que a soma da série com u_2 é dada por

$$u_2(x) = \frac{1}{2} \log \frac{1+x}{1-x} \quad \text{para } |x| < 1.$$

- (b) Verificar directamente que a função u_2 da alínea (a) é uma solução da equação de Legendre quando $\alpha = 0$.
2. Mostrar que a função f definida por

$$f(x) = 1 - \frac{x}{2} \log \frac{1+x}{1-x}$$

para $|x| < 1$ satisfaz à equação de Legendre (6.35) com $\alpha = 1$. Expressar esta função como uma combinação linear das soluções u_1 e u_2 dadas na equação (6.38) e (6.39).

3. A equação de Legendre (6.35) pode escrever-se na forma

$$[(x^2 - 1)y']' - \alpha(\alpha + 1)y = 0.$$

(a) Se a, b, c são constantes com $a > b$ e $4c + 1 > 0$, mostrar que a equação diferencial do tipo

$$[(x - a)(x - b)y']' - cy = 0$$

pode transformar-se numa equação de Legendre mediante uma mudança de variável da forma $x = At + B$, com $A > 0$. Determinar A e B em função de a e b .

(b) Utilizar o método seguido em (a) para transformar a equação

$$(x^2 - x)y'' + (2x - 1)y' - 2y = 0$$

numa equação de Legendre.

4. Achar duas soluções independentes, em forma de série de potências, da equação de Hermite

$$y'' - 2xy' + 2\alpha y = 0$$

num intervalo da forma $(-r, r)$. Mostrar que uma destas soluções é um polinómio quando α é inteiro não negativo.

5. Determinar uma solução, em forma de série de potências, para a equação diferencial

$$xy'' + (3 + x^3)y' + 3x^2y = 0$$

válida para todo o valor de x . Determinar uma segunda solução da forma $y = x^{-2} \sum a_n x^n$ válida para $x \neq 0$.

6. Determinar uma solução, em forma de série de potências, da equação diferencial

$$x^2 y'' + x^2 y' - (\alpha x + 2)y = 0$$

válida num intervalo da forma $(-r, r)$.

7. Dadas duas funções analíticas A e B num intervalo $(x_0 - r, x_0 + r)$, por exemplo

$$A(x) = \sum_{n=0}^{\infty} a_n (x - x_0)^n, \quad B(x) = \sum_{n=0}^{\infty} b_n (x - x_0)^n.$$

pode demonstrar-se que o produto $C(x) = A(x) B(x)$ é também uma função analítica em $(x_0 - r, x_0 + r)$. Este exercício mostra que C admite um desenvolvimento em série de potências

$$C(x) = \sum_{n=0}^{\infty} c_n (x - x_0)^n, \quad \text{onde } c_n = \sum_{k=0}^n a_k b_{n-k}.$$

(a) Utilizar a regra de Leibniz para a derivada de ordem n de um produto para demonstrar que a derivada de ordem n de C é dada por

$$C^{(n)}(x) = \sum_{k=0}^n \binom{n}{k} A^{(k)}(x) B^{(n-k)}(x).$$

(b) Fazer uso do facto de que $A^{(k)}(x_0) = k! a_k$ e $B^{(n-k)}(x_0) = (n-k)! b_{n-k}$ para obter

$$C^{(n)}(x_0) = n! \sum_{k=0}^n a_k b_{n-k}.$$

Visto que $C^{(n)}(x_0) = n! c_n$, isto prova a fórmula pretendida para c_n .

Nos exercícios 8 a 14, $P_n(x)$ representa o polinómio de Legendre de grau n . Estes exercícios constituem esquemas de demonstrações das propriedades dos polinómios de Legendre referidos na Secção 6.20.

8. (a) Recorrer à fórmula de Rodrigues para demonstrar que

$$P_n(x) = \frac{1}{2^n} (x+1)^n + (x-1)Q_n(x),$$

onde $Q_n(x)$ é um polinómio.

(b) Provar que $P_n(1) = 1$ e que $P_n(-1) = (-1)^n$.

(c) Provar que $P_n(x)$ é o único polinómio solução da equação de Legendre (com $\alpha = n$) tendo o valor 1 quando $x = 1$.

9. (a) Usar as equações diferenciais a que satisfazem P_n e P_m para mostrar que

$$[(1-x^2)(P_n P'_m - P'_n P_m)]' = [n(n+1) - m(m+1)]P_n P_m.$$

(b) Se $m \neq n$, integrar a equação da alínea (a) de -1 a 1 para estabelecer outra demonstração da relação de ortogonalidade

$$\int_{-1}^1 P_n(x) P_m(x) dx = 0.$$

10. (a) Seja $f(x) = (x^2 - 1)^n$. Recorrendo à integração por partes provar que

$$\int_{-1}^1 f^{(n)}(x) f^{(n)}(x) dx = - \int_{-1}^1 f^{(n+1)}(x) f^{(n-1)}(x) dx.$$

Aplicar esta fórmula repetidamente para deduzir que o integral do primeiro membro é igual a

$$2(2n)! \int_0^1 (1-x^2)^n dx.$$

(b) A substituição $x = \cos t$ transforma o integral $\int_0^1 (1-x^2)^n dx$ em $\int_0^{\pi/2} \sin^{2n+1} t \, dt$. Utilizar a relação

$$\int_0^{\pi/2} \sin^{2n+1} t \, dt = \frac{2n(2n-2) \cdots 2}{(2n+1)(2n-1) \cdots 3 \cdot 1}$$

e a fórmula de Rodrigues para obter

$$\int_{-1}^1 [P_n(x)]^2 dx = \frac{2}{2n+1}.$$

11. (a) Mostrar que

$$P_n(x) = \frac{(2n)!}{2^n(n!)^2} x^n + Q_n(x),$$

onde $Q_n(x)$ é um polinômio de grau inferior a n .

(b) Exprimir o polinômio $f(x) = x^4$ como uma combinação linear de P_0, P_1, P_2, P_3 e P_4 .

(c) Mostrar que todo o polinômio f de grau n se pode exprimir como uma combinação linear dos polinômios de Legendre P_0, P_1, \dots, P_n .

12. (a) Se f é um polinômio de grau n , escrever

$$f(x) = \sum_{k=0}^n c_k P_k(x).$$

[Isto é possível devido ao Exercício 11(c).] Para um valor de m fixo, $0 \leq m \leq n$, multiplicar ambos os membros desta equação por $P_m(x)$ e integrar de -1 a 1 . Utilizar os Exercícios 9(b) e 10(b) para deduzir a relação

$$c_m = \frac{2m+1}{2} \int_{-1}^1 f(x) P_m(x) dx.$$

13. Utilizar os Exercícios 9 e 11 para demonstrar que $\int_{-1}^1 g(x) P_n(x) dx = 0$ para todo polinômio g de grau inferior a n .

14. (a) Com o auxílio do teorema de Rolle provar que P_n não pode ter raízes múltiplas no intervalo aberto $(-1, 1)$. Por outras palavras, quaisquer raízes de P_n que estejam em $(-1, 1)$ serão raízes simples.

(b) Suponhamos que P_n tem m raízes no intervalo $(-1, 1)$. Se $m = 0$, seja $Q_0(x) = 1$. Se $m \geq 1$, seja

$$Q_m(x) = (x - x_1)(x - x_2) \cdots (x - x_m),$$

onde x_1, x_2, \dots, x_m são m raízes de P_n em $(-1, 1)$. Provar que, em cada ponto x de $(-1, 1)$, $Q_m(x)$ tem o mesmo sinal que $P_n(x)$.

(c) Com recurso à alínea (b) e ao Exercício 13, provar que a desigualdade $m < n$ conduz a uma conclusão absurda. Significa isto que P_n tem n raízes reais distintas, todas situadas no intervalo aberto $(-1, 1)$.

15. (a) Provar que o valor do integral $\int_{-1}^1 P_n(x) P'_{n+1}(x) dx$ é independente de n .

(b) Calcular o integral $\int_{-1}^1 x P_n(x) P'_{n-1}(x) dx$.

6.22. O método de Frobenius

Na Secção 6.17 aprendemos a determinar soluções, em forma de séries de potências, da equação diferencial

$$(6.43) \quad y'' + P_1(x)y' + P_2(x)y = 0$$

num intervalo centrado num ponto x_0 , onde os coeficientes P_1 e P_2 são funções analíticas. Se um dos coeficientes P_1 ou P_2 não for analítico numa vizinhança de x_0 , pode ou não existir a solução série de potências na vizinhança de x_0 . Por exemplo, suponhamos que tentamos determinar uma série de potências solução da equação diferencial

$$(6.44) \quad x^2 y'' - y' - y = 0$$

numa vizinhança de $x_0 = 0$. Se admitimos que existe uma solução da forma $y = \sum a_k x^k$ e substituímos esta série na equação diferencial somos conduzidos à formula

$$a_{n+1} = \frac{n^2 - n - 1}{n + 1} a_n.$$

Embora isto nos dê uma série de potências $y = \sum a_k x^k$ que verifica formalmente (6.44), o critério do cociente mostra que esta série de potências converge *unicamente* para $x = 0$. Assim, não existe uma solução de (6.44) sob a forma de uma série de potências, válida em qualquer intervalo aberto centrado na origem. Este exemplo não contradiz o teorema 6.13, porque quando escrevemos a equação (6.44) na forma (6.43) encontramos que os coeficientes P_1 e P_2 são dados por

$$P_1(x) = -\frac{1}{x^2} \quad \text{e} \quad P_2(x) = -\frac{1}{x^2}.$$

Estas funções não admitem desenvolvimentos em série de potências em torno da origem. A dificuldade aqui é que o coeficiente de y'' em (6.44) tem o valor 0 quando $x = 0$, por outras palavras, a equação diferencial tem um ponto singular em $x = 0$.

Para apreciar as dificuldades que se apresentam na investigação das equações diferenciais na vizinhança de um ponto singular é necessário um conhecimento da teoria das funções de um variável complexa. Contudo, alguns casos particulares importantes das equações com pontos singulares podem ser tratados por métodos elementares. Por exemplo, suponhamos a equação diferencial em (6.43) equivalente a uma equação diferencial da forma

$$(6.45) \quad (x - x_0)^2 y'' + (x - x_0)P(x)y' + Q(x)y = 0,$$

em que P e Q admitem desenvolvimentos em séries de potências em algum intervalo aberto $(x_0 - r, x_0 + r)$. Neste caso dizemos que x_0 é um ponto singular *regular* da equação. Se dividirmos ambos os membros de (6.45) por $(x - x_0)^2$ a equação vem

$$y'' + \frac{P(x)}{x - x_0} y' + \frac{Q(x)}{(x - x_0)^2} y = 0$$

para $x \neq x_0$. Se $P(x_0) \neq 0$ ou $Q(x_0) \neq 0$, ou se $Q(x_0) = 0$ e $Q'(x_0) \neq 0$, quer o coeficiente de y' quer o coeficiente de y não admitirão desenvolvimento em série de potências em torno do ponto x_0 , pelo que o teorema 6.13 não será aplicável. Em 1873 o matemático alemão Georg Frobenius (1849-1917) desenvolveu um método muito prático para tratar tais equações. Enunciaremos o teorema de Frobenius, mas não daremos a demonstração†. Na próxima Secção apresentaremos pormenores da demonstração para um caso especial importante, a equação de Bessel.

O teorema de Frobenius desdobra-se em duas partes, que dependem da natureza das raízes da equação quadrática

$$(6.46) \quad t(t-1) + P(x_0)t + Q(x_0) = 0.$$

Esta equação quadrática chama-se a *equação indicial* da equação diferencial dada (6.45). Os coeficientes $P(x_0)$ e $Q(x_0)$ são termos constantes no desenvolvimento de P e Q em série de potências. Sejam α_1 e α_2 raízes da equação indicial. Estas raízes podem ser reais ou complexas, iguais ou distintas. O tipo de solução obtido pelo método de Frobenius depende de se estas raízes diferem ou não por um número inteiro.

TEOREMA 6.14. PRIMEIRO CASO DO TEOREMA DE FROBENIUS. *Sejam α_1 e α_2 raízes da equação indicial e admita-se que $\alpha_1 - \alpha_2$ não é um número inteiro. A equação diferencial (6.45) tem duas soluções independentes u_1 e u_2 da forma*

$$(6.47) \quad u_1(x) = |x - x_0|^{\alpha_1} \sum_{n=0}^{\infty} a_n (x - x_0)^n, \quad \text{com } a_0 = 1,$$

e

$$(6.48) \quad u_2(x) = |x - x_0|^{\alpha_2} \sum_{n=0}^{\infty} b_n (x - x_0)^n, \quad \text{com } b_0 = 1.$$

Ambas as séries convergem no intervalo $|x - x_0| < r$, e a equação diferencial é satisfeita para $0 < |x - x_0| < r$.

TEOREMA 6.15. SEGUNDO CASO DO TEOREMA DE FROBENIUS. *Sejam α_1 e α_2 as raízes da equação indicial e admita-se que $\alpha_1 - \alpha_2 = N$, um inteiro não negativo. A equação diferencial (6.45) tem uma solução u_1 da forma (6.47) e outra solução independente u_2 da forma*

$$(6.49) \quad u_2(x) = |x - x_0|^{\alpha_2} \sum_{n=0}^{\infty} b_n (x - x_0)^n + C u_1(x) \log |x - x_0|,$$

† Para o estudo da demonstração ver E. Hille, *Analysis*, Vol. II, Blaisdell Pub. Co., 1966, ou E. A. Coddington, *An Introduction to Ordinary Differential Equations*, Prentice-Hall, 1961.

onde $b_0 = 1$. A constante C é diferente de zero se $N = 0$. Se $N > 0$, a constante C pode ou não ser zero. Como no caso anterior ambas as séries convergem no intervalo $|x - x_0| < r$ e as soluções são válidas por $0 < |x - x_0| < r$.

6.23. A equação de Bessel

Vamos aqui aplicar o método sugerido por Frobenius para resolver a equação de Bessel

$$(6.49) \quad x^2 y'' + xy' + (x^2 - \alpha^2)y = 0,$$

onde α é uma constante não negativa. Esta equação aparece em problemas relativos a vibrações de membranas, fluxo de calor em cilindros e propagação de corrente elétrica em conductores cilíndricos. Algumas das suas soluções são conhecidas por *funções de Bessel*. As funções de Bessel aparecem igualmente em certos problemas da Teoria dos Números. A equação recebeu o nome do astrónomo alemão F. W. Bessel (1784-1816), muito embora tivesse já aparecido anteriormente nas investigações de Daniel Bernoulli (1732) e Euler (1764).

A equação de Bessel tem a forma (6.45) com $x_0 = 0$, $P(x) = 1$, e $Q(x) = x^2 - \alpha^2$, pelo que o ponto x_0 é um ponto singular regular. Visto que P e Q são analíticas em todo o eixo real, vamos tentar determinar soluções da forma

$$(6.50) \quad y = |x|^t \sum_{n=0}^{\infty} a_n x^n,$$

com $a_0 \neq 0$, válida para todo real x , com a possível excepção de $x = 0$.

Consideramos, em primeiro lugar, $x > 0$, pelo que $|x|^t = x^t$. A derivação de (6.50) dá-nos

$$y' = tx^{t-1} \sum_{n=0}^{\infty} a_n x^n + x^t \sum_{n=0}^{\infty} n a_n x^{n-1} = x^{t-1} \sum_{n=0}^{\infty} (n+t) a_n x^n.$$

Analogamente, obtemos

$$y'' = x^{t-2} \sum_{n=0}^{\infty} (n+t)(n+t-1) a_n x^n.$$

Se $L(y) = x^2 y'' + xy' + (x^2 - \alpha^2)y$, obtemos

$$\begin{aligned} L(y) &= x^t \sum_{n=0}^{\infty} (n+t)(n+t-1) a_n x^n + x^t \sum_{n=0}^{\infty} (n+t) a_n x^n \\ &\quad + x^t \sum_{n=0}^{\infty} a_n x^{n+2} - x^t \sum_{n=0}^{\infty} \alpha^2 a_n x^n = x^t \sum_{n=0}^{\infty} [(n+t)^2 - \alpha^2] a_n x^n + x^t \sum_{n=0}^{\infty} a_n x^{n+2}. \end{aligned}$$

Façamos agora $L(y) = 0$, suprimamos x^t e tentemos determinar os a_n de modo que o coeficiente de cada potência de x seja nulo. Para o termo independente é necessário que $(t^2 - \alpha^2)a_0 = 0$. Como procuramos uma solução com $a_0 \neq 0$, isto exige que

$$(6.51) \quad t^2 - \alpha^2 = 0.$$

Esta é a equação indicial. As suas raízes α e $-\alpha$ são os únicos valores possíveis de t que podem dar-nos uma solução do tipo desejado.

Consideremos em primeiro lugar a escolha $t = \alpha$. Para este t , as restantes equações para a determinação dos coeficientes escrevem-se

$$(6.52) \quad [(1 + \alpha)^2 - \alpha^2]a_1 = 0 \quad \text{e} \quad [(n + \alpha)^2 - \alpha^2]a_n + a_{n-2} = 0$$

para $n \geq 2$. Uma vez que $\alpha \geq 0$, a primeira destas equações implica que $a_1 = 0$. A segunda fórmula pode escrever-se

$$(6.53) \quad a_n = -\frac{a_{n-2}}{(n + \alpha)^2 - \alpha^2} = -\frac{a_{n-2}}{n(n + 2\alpha)},$$

pelo que $a_3 = a_5 = a_7 = \dots = 0$. Para os coeficientes com índices pares temos

$$a_2 = \frac{-a_0}{2(2 + 2\alpha)} = \frac{-a_0}{2^2(1 + \alpha)}, \quad a_4 = \frac{-a_2}{4(4 + 2\alpha)} = \frac{(-1)^2 a_0}{2^4 2! (1 + \alpha)(2 + \alpha)},$$

$$a_6 = \frac{-a_4}{6(6 + 2\alpha)} = \frac{(-1)^3 a_0}{2^6 3! (1 + \alpha)(2 + \alpha)(3 + \alpha)},$$

e, em geral,

$$a_{2n} = \frac{(-1)^n a_0}{2^{2n} n! (1 + \alpha)(2 + \alpha) \cdots (n + \alpha)}.$$

Consequentemente a escolha $t = \alpha$ dá-nos a solução

$$y = a_0 x^\alpha \left(1 + \sum_{n=1}^{\infty} \frac{(-1)^n x^{2n}}{2^{2n} n! (1 + \alpha)(2 + \alpha) \cdots (n + \alpha)} \right)$$

O critério do cociente mostra que a série de potências que aparece nesta fórmula converge para todo o real x .

Nesta discussão admitimos que $x > 0$. Se $x < 0$ podemos repetir a discussão com x^t substituído por $(-x)^t$. Encontramos ainda que t deve satisfazer à equação $t^2 - \alpha^2 = 0$. Fazendo $t = \alpha$ obtemos a mesma solução, excepto que o factor x^α é substituído por $(-x)^\alpha$. Deste modo a função f_α dada pela equação

$$(6.54) \quad f_\alpha(x) = a_0 |x|^\alpha \left(1 + \sum_{n=1}^{\infty} \frac{(-1)^n x^{2n}}{2^{2n} n! (1 + \alpha)(2 + \alpha) \cdots (n + \alpha)} \right)$$

é uma solução da equação de Bessel válida para todo o real $x \neq 0$. Para os valores de α para os quais existem $f'_\alpha(0)$ e $f''_\alpha(0)$, a solução também é válida para $x = 0$.

Consideremos agora a raiz $t = -\alpha$ da equação indicial. Obtemos, em vez de (6.52), as equações

$$[(1 - \alpha)^2 - \alpha^2]a_1 = 0 \quad \text{e} \quad [(n - \alpha)^2 - \alpha^2]a_n + a_{n-2} = 0,$$

que são equivalentes a

$$(1 - 2\alpha)a_1 = 0 \quad \text{e} \quad n(n - 2\alpha)a_n + a_{n-2} = 0.$$

Se 2α não for inteiro, estas equações dão-nos $a_1 = 0$ e

$$a_n = -\frac{a_{n-2}}{n(n - 2\alpha)}$$

para $n \geq 2$. Uma vez que esta fórmula de recorrência é a mesma que (6.53), com α substituído por $-\alpha$, somos conduzidos à solução

$$(6.55) \quad f_{-\alpha}(x) = a_0 |x|^{-\alpha} \left(1 + \sum_{n=1}^{\infty} \frac{(-1)^n x^{2n}}{2^{2n} n! (1 - \alpha)(2 - \alpha) \cdots (n - \alpha)} \right)$$

válida para todo o real $x \neq 0$.

A solução $f_{-\alpha}$ foi obtida sob a hipótese de que 2α não é um inteiro positivo. Contudo, a série para $f_{-\alpha}$ tem ainda sentido no caso de 2α ser um inteiro positivo, com tanto que α o não seja. Pode verificar-se que $f_{-\alpha}$ satisfaz à equação de Bessel para tal valor de α . Assim, para cada $\alpha \geq 0$, temos a série solução f_{α} dada pela equação (6.54); e se α não for um inteiro não negativo encontrámos outra solução $f_{-\alpha}$ dada pela equação (6.55). As duas soluções f_{α} e $f_{-\alpha}$ são independentes, uma vez que uma delas $\rightarrow \infty$ quando $x \rightarrow 0$, e a outra não. Seguidamente simplificaremos a forma das soluções. Para o conseguir necessitamos de algumas propriedades da função gama de Euler que passamos a recordar.

Para cada real $s > 0$ define-se $\Gamma(s)$ pelo integral impróprio

$$\Gamma(s) = \int_{0+}^{\infty} t^{s-1} e^{-t} dt.$$

Este integral converge se $s > 0$ e diverge se $s \leq 0$. A integração por partes conduz à equação funcional

$$(6.56) \quad \Gamma(s + 1) = s \Gamma(s).$$

Isto implica que

$$\Gamma(s+2) = (s+1)\Gamma(s+1) = (s+1)s\Gamma(s),$$

$$\Gamma(s+3) = (s+2)\Gamma(s+2) = (s+2)(s+1)s\Gamma(s),$$

e, em geral,

$$(6.57) \quad \Gamma(s+n) = (s+n-1) \cdots (s+1)s\Gamma(s)$$

para todo o inteiro positivo n . Visto que $\Gamma(1) = \int_0^\infty e^{-t} dt = 1$, quando fazemos $s = 1$ em (6.57) encontramos

$$\Gamma(n+1) = n!.$$

Assim se conclui que a função gama é uma extensão, para os números reais, positivos da função factorial definida para números inteiros.

A equação funcional (6.56) pode usar-se para a extensão de $\Gamma(s)$ a valores negativos de s , não inteiros. Escrevemos (6.56) na forma

$$(6.58) \quad \Gamma(s) = \frac{\Gamma(s+1)}{s}.$$

O segundo membro tem sentido se $s+1 > 0$ e $s \neq 0$. Deste modo, podemos usar esta equação para definir $\Gamma(s)$ se $-1 < s < 0$. O segundo membro de (6.58) tem agora significado se $s+2 > 0$, $s \neq -1$, $s \neq 0$, e podemos usar esta equação para definir $\Gamma(s)$ para $-2 < s < -1$. Continuando deste modo, podemos generalizar a definição de $\Gamma(s)$ por indução ao intervalo aberto da forma $-n < s < -n+1$, em que n é um inteiro positivo qualquer. A equação funcional (6.56) e a sua generalização (6.57) são agora válidas para todo o real s para o qual ambos os membros tenham significado.

Voltemos de novo à discussão da equação de Bessel. A série para f_α na equação (6.54) contém o produto $(1+\alpha)(2+\alpha)\cdots(n+\alpha)$. Podemos exprimir este produto por intermédio da função gama fazendo $s = 1 + \alpha$ em (6.57). Obtemos

$$(1+\alpha)(2+\alpha)\cdots(n+\alpha) = \frac{\Gamma(n+1+\alpha)}{\Gamma(1+\alpha)}.$$

Deste modo, se escolhermos $a_0 = 2^{-\alpha}/\Gamma(1+\alpha)$ na equação (6.54) e representamos a função resultante $f_\alpha(x)$ por $J_\alpha(x)$ quando $x > 0$, a solução para $x > 0$ pode escrever-se

$$(6.59) \quad J_\alpha(x) = \left(\frac{x}{2}\right)^\alpha \sum_{n=0}^{\infty} \frac{(-1)^n}{n! \Gamma(n+1+\alpha)} \left(\frac{x}{2}\right)^{2n}.$$

A função J_α definida por esta equação para $x > 0$ e $\alpha \geq 0$ chama-se a *função de Bessel de primeira espécie e de ordem α* . Quando α é um inteiro não negativo por exemplo $\alpha = p$, a função de Bessel é dada pela série de potências

$$J_p(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n! (n+p)!} \left(\frac{x}{2}\right)^{2n+p} \quad (p = 0, 1, 2, \dots).$$

Esta é também uma solução da equação de Bessel para $x < 0$. Extensas tabelas das funções de Bessel estão já calculadas. Na figura 6.2 representam-se os gráficos de duas funções, J_0 e J_1 .

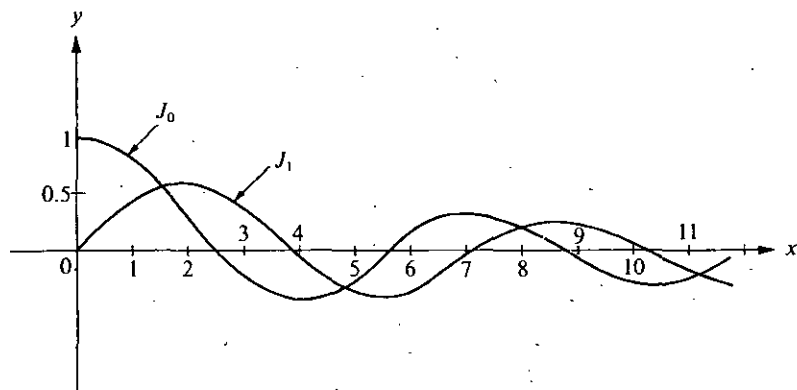


FIG. 6.2. Gráficos das funções de Bessel J_0 e J_1 .

Podemos definir uma nova função $J_{-\alpha}$ por substituição de α por $-\alpha$ em (6.59), se α é tal que $\Gamma(n+1-\alpha)$ tem significado; isto é, se α não for um inteiro positivo. Portanto, se $x > 0$ e $\alpha > 0$, $\alpha \neq 1, 2, 3, \dots$, definimos

$$J_{-\alpha}(x) = \left(\frac{x}{2}\right)^{-\alpha} \sum_{n=0}^{\infty} \frac{(-1)^n}{n! \Gamma(n+1-\alpha)} \left(\frac{x}{2}\right)^{2n}.$$

Fazendo $s = 1 - \alpha$ em (6.57) obtemos

$$\Gamma(n+1-\alpha) = (1-\alpha)(2-\alpha) \cdots (n-\alpha) \Gamma(1-\alpha)$$

e vemos que a série para $J_{-\alpha}(x)$ é a mesma que para $J_{\alpha}(x)$ na equação (6.55) com $a_0 = 2^{\alpha} \Gamma(1-\alpha)$, $x > 0$. Portanto, se α não for um inteiro positivo, $J_{-\alpha}$ é uma solução da equação de Bessel para $x > 0$.

Se α não for inteiro, as duas soluções $J_{\alpha}(x)$ e $J_{-\alpha}(x)$ são linearmente independentes no semi eixo real positivo (visto que o seu cociente não é constante) e a solução geral da equação de Bessel para $x > 0$ é

$$y = c_1 J_{\alpha}(x) + c_2 J_{-\alpha}(x).$$

Se α é um inteiro não negativo, por exemplo $\alpha = p$, encontrámos unicamente a solução J_p e os seus produtos por constantes válidos para $x > 0$. Outra solução, indepen-

dente desta, pode ser encontrada pelo método descrito no Exercício 4 da Secção 6.16. Este estabelece que se u_1 é uma solução de $y'' + P_1 y' + P_2 y = 0$ que nunca se anula num intervalo I , uma segunda solução u_2 independente de u_1 é dada pelo integral

$$u_2(x) = u_1(x) \int_c^x \frac{Q(t)}{[u_1(t)]^2} dt,$$

onde $Q(x) = e^{-\int P_1(x) dx}$. Para a equação de Bessel temos $P_1(x) = 1/x$, pelo que $Q(x) = 1/x$ e uma segunda solução u_2 é dada pela fórmula

$$(6.60) \quad u_2(x) = J_p(x) \int_c^x \frac{1}{t[J_p(t)]^2} dt,$$

se c e x estão num intervalo I no qual J_p não se anula.

A esta segunda solução pode dar-se-lhe outras formas. Por exemplo, da equação (6.59) podemos escrever

$$\frac{1}{[J_p(t)]^2} = \frac{1}{t^{2p}} g_p(t),$$

onde $g_p(0) \neq 0$. No intervalo I a função g_p admite um desenvolvimento em série de potências

$$g_p(t) = \sum_{n=0}^{\infty} A_n t^n$$

o qual pode determinar-se igualando coeficiente na identidade $g_p(t)[J_p(t)]^2 = t^p$. Se admitirmos a existência de um tal desenvolvimento, o integrando em (6.60) toma a forma

$$\frac{1}{t[J_p(t)]^2} = \frac{1}{t^{2p+1}} \sum_{n=0}^{\infty} A_n t^n.$$

Integrando termo a termo de c a x obtemos um termo logarítmico $A_{2p} \log x$ (da potência t^{-1}), mais uma série da forma $x^{-2p} \sum B_n x^n$. Por conseguinte a equação (6.60) toma a forma

$$u_2(x) = A_{2p} J_p(x) \log x + J_p(x) x^{-2p} \sum_{n=0}^{\infty} B_n x^n.$$

Pode demonstrar-se que o coeficiente $A_{2p} \neq 0$. Se multiplicamos $u_2(x)$ por $1/A_{2p}$ a solução resultante designa-se por $K_p(x)$ e tem a forma

$$K_p(x) = J_p(x) \log x + x^{-p} \sum_{n=0}^{\infty} C_n x^n.$$

Esta é a forma da solução indicada pelo segundo caso do teorema de Frobenius.

Uma vez chegados a esta fórmula, podemos verificar que existe realmente uma solução deste tipo, substituindo o segundo membro na equação de Bessel, e determinando os coeficientes C_n de modo que a equação seja satisfeita. Os pormenores deste cálculo são extensos e por tal motivo são omitidos. O resultado final pode ser expresso por

$$K_p(x) = J_p(x) \log x - \frac{1}{2} \left(\frac{x}{2}\right)^{-p} \sum_{n=0}^{p-1} \frac{(p-n-1)!}{n!} \left(\frac{x}{2}\right)^{2n} - \frac{1}{2} \left(\frac{x}{2}\right)^p \sum_{n=0}^{\infty} (-1)^n \frac{h_n + h_{n+p}}{n!(n+p)!} \left(\frac{x}{2}\right)^{2n},$$

onde $h_0 = 0$ e $h_n = 1 + \frac{1}{2} + \dots + 1/n$ para $n \geq 1$. A série do segundo membro converge para todo o real x . A função K_p definida para $x > 0$ por esta fórmula chama-se a *função de Bessel de segunda espécie de ordem p* . Porque K_p não é o produto de uma constante por J_p , a solução geral da equação de Bessel neste caso para $x > 0$ é

$$y = c_1 J_p(x) + c_2 K_p(x).$$

No conjunto de exercícios que se apresenta a seguir são analisadas mais algumas propriedades das funções de Bessel.

6.24. Exercícios

1. (a) Seja f uma solução qualquer da equação de Bessel de ordem α e $g(x) = x^{1/2} f(x)$ para $x > 0$. Demonstrar que g verifica a equação diferencial

$$y'' + \left(1 + \frac{1 - 4\alpha^2}{4x^2}\right)y = 0.$$

(b) Quando $4\alpha^2 = 1$ a equação diferencial da alínea (a) transformase em $y'' + y = 0$; a sua solução geral é $y = A \cos x + B \sin x$. Utilizar este facto e a igualdade $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ para mostrar que, para $x > 0$,

$$J_{1/2}(x) = \left(\frac{2}{\pi x}\right)^{1/2} \sin x \quad \text{e} \quad J_{-1/2}(x) = \left(\frac{2}{\pi x}\right)^{1/2} \cos x.$$

- (c) Deduzir as fórmulas da alínea (b) directamente a partir das séries para $J_{1/2}(x)$ e $J_{-1/2}(x)$.
2. Utilizar a representação das funções de Bessel em séries de potências para provar que

$$(a) \quad \frac{d}{dx} (x^\alpha J_\alpha(x)) = x^\alpha J_{\alpha-1}(x),$$

† A mudança de variável $t = u^2$ dá-nos

$$\Gamma(\frac{1}{2}) = \int_0^\infty t^{-1/2} e^{-t} dt = 2 \int_0^\infty e^{-u^2} du = \sqrt{\pi}.$$

(Ver Exercício 16 da Secção 11.28 para a demonstração de que $2 \int_0^\infty e^{-u^2} du = \sqrt{\pi}$.)

$$(b) \frac{d}{dx} (x^{-\alpha} J_{\alpha}(x)) = -x^{-\alpha} J_{\alpha+1}(x).$$

3. Seja $F_n(x) = x^n J_n(x)$ e $G_n(x) = x^{-n} J_n(x)$ para $x > 0$. Observar que cada raiz positiva de J_n é raiz de F_n e é também uma raiz de G_n . Utilizar o teorema de Rolle e o Exercício 2 para demonstrar que os raízes positivos de J_n e J_{n+1} se intercalam, isto é, existe uma raiz de J_n entre cada par de raízes positivas de J_{n+1} , e uma raiz de J_{n+1} entre cada par de raízes positivas de J_n . (Ver figura 6.2).
4. (a) A partir das relações do Exercício 2 deduzir as fórmulas de recorrência

$$\frac{\alpha}{x} J_{\alpha}(x) + J'_{\alpha}(x) = J_{\alpha-1}(x) \quad \text{e} \quad \frac{\alpha}{x} J_{\alpha}(x) - J'_{\alpha}(x) = J_{\alpha+1}(x).$$

- (b) Utilizar as relações da alínea (a) para deduzir as fórmulas

$$J_{\alpha-1}(x) + J_{\alpha+1}(x) = \frac{2\alpha}{x} J_{\alpha}(x) \quad \text{e} \quad J_{\alpha-1}(x) - J_{\alpha+1}(x) = 2J'_{\alpha}(x).$$

5. Do exercício 1(b) e de uma fórmula de recorrência conveniente mostrar que

$$J_{3/2}(x) = \left(\frac{2}{\pi x}\right)^{1/2} \left(\frac{\sin x}{x} - \cos x\right).$$

Achar uma forma análoga para $J_{-3/2}(x)$. *Observação.* $J_{\alpha}(x)$ é uma função elementar para todo α que seja igual a um meio de um inteiro ímpar.

6. Provar que

$$\frac{1}{2} \frac{d}{dx} (J_{\alpha}^2(x) + J_{\alpha+1}^2(x)) = \frac{\alpha}{x} J_{\alpha}^2(x) - \frac{\alpha+1}{x} J_{\alpha+1}^2(x)$$

e

$$\frac{d}{dx} (x J_{\alpha}(x) J_{\alpha+1}(x)) = x (J_{\alpha}^2(x) - J_{\alpha+1}^2(x)).$$

7. (a) Usar as identidades do Exercício 6 para provar que

$$J_0^2(x) + 2 \sum_{n=1}^{\infty} J_n^2(x) = 1 \quad \text{e} \quad \sum_{n=0}^{\infty} (2n+1) J_n(x) J_{n+1}(x) = \frac{1}{2} x.$$

(b) A partir da alínea (a), demonstrar que $|J_0(x)| \leq 1$ e $|J_n(x)| \leq \frac{1}{2} \sqrt{2}$ para $n = 1, 2, 3, \dots$ e todo $x \leq 0$.

8. Seja $g_n(x) = x^{1/2} J_n(ax^b)$ para $x > 0$, onde a e b são constantes não nulas. Mostrar que g_n satisfaz à equação diferencial

$$x^2 y'' + (a^2 b^2 x^{2b} + \frac{1}{4} - \alpha^2 b^2) y = 0$$

se, e só se, f_{α} é uma solução da equação de Bessel de ordem α .

9. Recorrer ao Exercício 8 para exprimir a solução geral de cada uma das seguintes equações diferenciais por intermédio das funções de Bessel para $x > 0$.

- (a) $y'' + xy = 0$. (c) $y'' + x^m y = 0$.
 (b) $y'' + x^2 y = 0$. (d) $x^2 y'' + (x^4 + \frac{1}{8})y = 0$.
10. Generalizar o Exercício 8 quando f_a e g_a estão relacionados pela equação $g_a(x) = x^a f_a(ax^b)$ para $x > 0$. Achar então a solução geral das seguintes equações por intermédio das funções de Bessel para $x > 0$.
- (a) $xy'' + 6y' + y = 0$. (c) $xy'' + 6y' + x^4 y = 0$.
 (b) $xy'' + 6y' + xy = 0$. (d) $x^2 y'' - xy' + (x + 1)y = 0$.
11. Existe uma identidade para a função de Bessel da forma

$$J_2(x) - J_0(x) = aJ_c'(x),$$

com a e c constantes. Determinar a e c .

12. Determinar uma série de potências solução da equação diferencial $xy'' + y' + y = 0$ convergente para $-\infty < x < \infty$. Mostrar que para $x > 0$ pode exprimir-se por intermédio de uma função de Bessel.
13. Considere uma equação diferencial linear de segunda ordem da forma

$$x^2 A(x)y'' + xP(x)y' + Q(x)y = 0,$$

onde $A(x)$, $P(x)$, e $Q(x)$ admitem desenvolvimentos em série de potências

$$A(x) = \sum_{k=0}^{\infty} a_k x^k, \quad P(x) = \sum_{k=0}^{\infty} p_k x^k, \quad Q(x) = \sum_{k=0}^{\infty} q_k x^k,$$

com $a_0 \neq 0$, convergente cada uma num intervalo aberto $(-r, r)$. Se a equação diferencial tem uma solução sob a forma de série da forma

$$y = x^t \sum_{n=0}^{\infty} c_n x^n,$$

válida para $0 < x < r$, mostrar que t satisfaz à equação quadrática da forma $t^2 + bt + c = 0$, e determine b e c em função dos coeficientes das séries de $A(x)$, $P(x)$ e $Q(x)$.

14. Considerar um caso particular do Exercício 13 no qual $A(x) = 1 - x$, $P(x) = \frac{1}{2}$ e $Q(x) = -\frac{1}{4}x$. Determinar uma série solução com t não inteiro.
15. A equação diferencial $2x^2 y'' + (x^2 - x)y' + y = 0$ tem duas soluções independentes da forma

$$y = x^t \sum_{n=0}^{\infty} c_n x^n,$$

válidas para $x > 0$. Determinar essas soluções.

16. A equação diferencial não linear $y'' + y + \alpha y^2 = 0$ é "ligeiramente" não linear se α for uma constante pequena, não nula. Suponha-se que existe uma solução que pode exprimir-se como uma série de potências de α da forma

$$y = \sum_{n=0}^{\infty} u_n(x) \alpha^n \quad (\text{válida em algum intervalo } 0 < \alpha < r)$$

e que esta solução satisfaz às condições iniciais $y = 1$ e $y' = 0$ quando $x = 0$. Para se ajustar a estas condições iniciais, tentamos escolher os coeficientes $u_n(x)$ de tal maneira que $u_0(0) = 1$, $u_0'(0) = 0$ e $u_n(0) = u_n'(0)$ para $n \geq 1$. Substituir esta série na equação diferencial, igualar as potências de α e determinar $u_0(x)$ e $u_1(x)$.

7

SISTEMAS DE EQUAÇÕES DIFERENCIAIS

7.1. Introdução

Embora o estudo das equações diferenciais tivesse começado no século XVII, só no século XIX os matemáticos se aperceberam de que relativamente poucas equações diferenciais podiam ser resolvidas por meios elementares. Os trabalhos de Cauchy, Liouville, e outros, mostraram a importância do estabelecimento de teoremas gerais garantindo a existência de soluções para certas classes específicas de equações diferenciais. O capítulo 6 indica o uso que pode fazer-se de um teorema de existência e unicidade no estudo de equações diferenciais lineares. Este capítulo é dedicado à demonstração daquele teorema e tópicos com ele relacionados.

A teoria da existência de equações diferenciais de ordem superior pode relacionar-se ao caso de primeira ordem pela introdução de *sistemas* de equações. Por exemplo, a equação de segunda ordem

$$(7.1) \quad y'' + 2ty' - y = e^t$$

pode transformar-se num sistema de duas equações de primeira ordem pela introdução de duas funções desconhecidas y_1 e y_2 , onde

$$y_1 = y, \quad y_2 = y'_1.$$

Então tem-se $y'_2 = y''_1 = y''$, e (7.1) pode escrever-se como um sistema de duas equações de primeira ordem:

$$(7.2) \quad \begin{aligned} y'_1 &= y_2 \\ y'_2 &= y_1 - 2ty_2 + e^t. \end{aligned}$$

Não é possível resolver estas equações separadamente pelos métodos do Capítulo 6, porque cada uma delas contém as duas funções desconhecidas.

Neste capítulo consideramos sistemas formados por n equações diferenciais lineares de primeira ordem contendo n funções desconhecidas y_1, \dots, y_n . Estes sistemas têm a forma

$$\begin{aligned} y_1' &= p_{11}(t)y_1 + p_{12}(t)y_2 + \dots + p_{1n}(t)y_n + q_1(t) \\ &\vdots \\ y_n' &= p_{n1}(t)y_1 + p_{n2}(t)y_2 + \dots + p_{nn}(t)y_n + q_n(t). \end{aligned} \quad (7.3)$$

As funções p_{ik} e q_i que aparecem em (7.3) consideram-se como funções dadas definidas num determinado intervalo J . As funções y_1, \dots, y_n são funções incógnitas que interessa determinar. Sistemas deste tipo chamam-se *sistemas lineares de primeira ordem*. Em geral, cada equação do sistema contém mais do que uma função incógnita pelo que as equações não se podem resolver separadamente.

Uma equação diferencial linear de ordem n pode sempre transformar-se num sistema linear. Suponhamos que a equação de ordem n é

$$y^{(n)} + a_1 y^{(n-1)} + \dots + a_n y = R(t), \quad (7.4)$$

com os coeficientes a_i funções dadas. Para transformar esta equação num sistema, escrevemos $y_1 = y$ e introduzimos uma nova função desconhecida para cada uma das derivadas sucessivas de y , isto é, escrevemos

$$y_1 = y, \quad y_2 = y_1', \quad y_3 = y_2', \dots, y_n = y_{n-1}',$$

e escrevemos (7.4) na forma de sistema

$$\begin{aligned} y_1' &= y_2 \\ y_2' &= y_3 \\ &\vdots \\ y_{n-1}' &= y_n \\ y_n' &= -a_n y_1 - a_{n-1} y_2 - \dots - a_1 y_n + R(t). \end{aligned} \quad (7.5)$$

A discussão dos sistemas pode ser consideravelmente simplificada pelo recurso à notação vectorial e matricial. Consideremos o sistema geral (7.3) e introduzamos as funções vectoriais $Y = (y_1, \dots, y_n)$, $Q = (q_1, \dots, q_n)$, e uma função matricial $P = [p_{ij}]$, definida pelas equações

$$Y(t) = (y_1(t), \dots, y_n(t)), \quad Q(t) = (q_1(t), \dots, q_n(t)), \quad P(t) = [p_{ji}(t)]$$

para cada t de J . Consideremos os vectores como matrizes colunas $n \times 1$ e escrevemos o sistema (7.3) na forma mais simples:

$$(7.6) \quad Y' = P(t)Y + Q(t).$$

Por exemplo, no sistema (7.2) temos

$$Y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad P(t) = \begin{bmatrix} 0 & 1 \\ 1 & -2t \end{bmatrix}, \quad Q(t) = \begin{bmatrix} 0 \\ e^t \end{bmatrix}.$$

No sistema (7.5) temos

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad P(t) = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{bmatrix}, \quad Q(t) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ R(t) \end{bmatrix}.$$

Um problema de valores iniciais para o sistema (7.6) é encontrar uma função vectorial Y , a qual satisfaz a (7.6) e que também satisfaz a uma condição inicial da forma $Y(a) = B$, onde $a \in J$ e $B = (b_1, b_2, \dots, b_n)$ é um vector n -dimensional dado.

No caso $n = 1$ (o caso escalar) sabemos do teorema 6.1 que, se P e Q são contínuas em J , todas as soluções de (7.6) são dados pela fórmula explícita

$$(7.7) \quad Y(x) = e^{A(x)}Y(a) + e^{A(x)} \int_a^x e^{-A(t)}Q(t) dt,$$

onde $A(x) = \int_a^x P(t)dt$, e a é qualquer ponto de J . Mostraremos que esta fórmula pode generalizar-se de maneira adequada para sistemas, isto é, quando $P(t)$ é uma função matricial $n \times n$ e $Q(t)$ é uma função vectorial n -dimensional. Para isso temos que atribuir um significado aos integrais de matrizes e às exponenciais de matrizes. Por tal motivo vamos efectuar breve digressão referente ao cálculo com funções matriciais.

7.2. Conceitos do cálculo para funções matriciais

A generalização dos conceitos de integral e derivada para funções matriciais é imediata. Se $P(t) = [p_{ij}(t)]$, definimos o integral $\int_a^b P(t) dt$ pela equação

$$\int_a^b P(t) dt = \left[\int_a^b p_{ij}(t) dt \right].$$

Quer dizer, o integral da matriz $P(t)$ é a matriz obtida por integração de cada elemento de $P(t)$, supondo evidentemente que cada elemento é integrável em $[a, b]$. O leitor pode verificar que a propriedade de linearidade para integrais se generaliza às funções matriciais.

Continuidade e derivabilidade de funções matriciais definem-se igualmente a partir dos respectivos elementos. Dizemos que uma função matricial $P = [p_{ij}]$ é contínua em t se cada elemento p_{ij} é contínuo em t . A derivada P' define-se pela matriz que se obtém derivando cada elemento,

$$P'(t) = [p'_{ij}(t)],$$

sempre que todas as derivadas $p'_{ij}(t)$ existam. É fácil provar as regras fundamentais da derivação para somas e produtos. Por exemplo, se P e Q são funções matriciais deriváveis, temos

$$(P + Q)' = P' + Q'$$

caso P e Q sejam do mesmo tipo e tem-se igualmente.

$$(PQ)' = PQ' + P'Q$$

se se define o produto PQ . A regra de derivação da função composta continua válida, isto é, se $F(t) = P[g(t)]$, com P uma função matricial derivável e g uma função escalar derivável, então $F'(t) = g'(t)P'[g(t)]$. O teorema da derivada nula, e o primeiro e segundo teoremas fundamentais do cálculo são também válidos para funções matriciais. As demonstrações destas propriedades são pedidas no conjunto de exercícios apresentados em 7.4.

A definição da exponencial de uma matriz não é tão simples e exige mais preparação. Tal conceito expõe-se na secção seguinte.

7.3. Séries de matrizes. Normas de matrizes

Seja $A = [a_{ij}]$ uma matriz $n \times n$ de elementos reais ou complexos. Pretendemos definir a exponencial e^A de maneira que possua algumas das propriedades fundamentais da exponencial ordinária de valores reais ou complexos. Em particular, pretende-se a verificação da propriedade

$$(7.8) \quad e^{tA} e^{sA} = e^{(t+s)A} \quad \text{para todo o par } s \text{ e } t,$$

reais e da relação

$$(7.9) \quad e^O = I,$$

onde O e I são as matrizes, $n \times n$, zero e identidade respectivamente. Pode parecer natural definir e^A como a matriz $[e^{a_{ij}}]$. Contudo, isso é inaceitável visto que tal definição não satisfaz a nenhuma das propriedades (7.8) e (7.9). Pelo contrário, definiremos e^A por meio de um desenvolvimento em série de potências

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}.$$

Sabemos que esta fórmula é válida se A for um número real ou complexo e provaremos que ela implica as propriedades (7.8) e (7.9) se A for uma matriz. Antes porém de poder-nos efectuar as demonstrações torna-se necessário explicar o que se entende por uma série convergente de matrizes.

DEFINIÇÃO DE UMA SÉRIE CONVERGENTE DE MATRIZES. *Dada uma sucessão $\{C_k\}$ de matrizes $m \times n$ cujos elementos são números reais ou complexos, designe-se o elemento ij de C_k por $c_{ij}^{(k)}$. Se todas as séries*

$$(7.10) \quad \sum_{k=1}^{\infty} c_{ij}^{(k)} \quad (i = 1, \dots, m; j = 1, \dots, n)$$

são convergentes, dizemos então que a série de matrizes $\sum_{k=1}^{\infty} C_k$ é convergente, e a sua soma define-se como a matriz $m \times n$ cujo elemento ij é a série (7.10).

Um critério de convergência simples e útil para uma série de matrizes pode ser estabelecido por intermédio da *norma* de uma matriz, uma generalização de valor absoluto de um número.

DEFINIÇÃO DE NORMA DE UMA MATRIZ. *Se $A = [a_{ij}]$ é uma matriz $m \times n$, de elementos reais ou complexos, a norma de A , representada por $\|A\|$, define-se como sendo o número não negativo dado por*

$$(7.11) \quad \|A\| = \sum_{i=1}^m \sum_{j=1}^n |a_{ij}|.$$

Por outras palavras, temos que a norma de A é a soma dos valores absolutos de todos os seus elementos. Existem outras definições de norma também usadas às vezes, mas aqui escolheu-se esta devido à facilidade com que, a partir dela, se podem demonstrar as propriedades seguintes.

TEOREMA 7.1. PROPRIEDADES FUNDAMENTAIS DAS NORMAS. *Para as matrizes retangulares A e B e todo o escalar real ou complexo c tem-se*

$$\|A + B\| \leq \|A\| + \|B\|, \|AB\| \leq \|A\| \|B\|, \|cA\| = |c| \|A\|.$$

Demonstração. Vamos demonstrar unicamente o resultado para $\|AB\|$, admitindo que A é $m \times n$ e B é $n \times p$. As outras demonstrações são mais simples e como tal são deixadas ao leitor como exercício.

Escrevendo $A = [a_{ik}]$, $B = [b_{kj}]$, temos $AB = [\sum_{k=1}^n a_{ik} b_{kj}]$, pelo que de (7.11) resulta

$$\|AB\| = \sum_{i=1}^m \sum_{j=1}^p \left| \sum_{k=1}^n a_{ik} b_{kj} \right| \leq \sum_{i=1}^m \sum_{k=1}^n |a_{ik}| \sum_{j=1}^p |b_{kj}| \leq \sum_{i=1}^m \sum_{k=1}^n |a_{ik}| \|B\| = \|A\| \|B\|.$$

Observe-se que no caso particular em que $B = A$, a desigualdade para $\|AB\|$ se converte em $\|A^2\| \leq \|A\|^2$. Por indução temos

$$\|A^k\| \leq \|A\|^k \quad \text{para } k = 1, 2, 3, \dots$$

Estas desigualdades são úteis na discussão da exponencial de uma matriz.

O teorema seguinte dá-nos uma útil condição suficiente para a convergência de uma série de matrizes.

TEOREMA 7.2. CRITÉRIO DE CONVERGÊNCIA DE UMA SÉRIE DE MATRIZES. *Se $\{C_k\}$ é uma sucessão de matrizes $m \times n$ tais que $\sum_{k=1}^{\infty} \|C_k\|$ converge, então a série de matrizes $\sum_{k=1}^{\infty} C_k$ também converge.*

Demonstração. Representemos por $c_{ij}^{(k)}$ o elemento ij de C_k . Visto que $|c_{ij}^{(k)}| \leq \|C_k\|$, a convergência de $\sum_{k=1}^{\infty} \|C_k\|$ implica a convergência absoluta de cada uma das séries $\sum_{k=1}^{\infty} c_{ij}^{(k)}$. Logo cada uma das séries $\sum_{k=1}^{\infty} c_{ij}^{(k)}$ é convergente, pelo que a série de matrizes $\sum_{k=1}^{\infty} C_k$ é convergente.

7.4. Exercícios

1. Verificar que a propriedade de linearidade dos integrais também é válida para integrais de funções matriciais.
2. Verificar cada uma das seguintes regras de derivação para funções matriciais, admitindo que P e Q são deriváveis. Em (a), P e Q devem ser do mesmo tipo, de modo a que $P + Q$ tenha significado. Em (b) e (d) não é necessários que sejam do mesmo tipo, bastando que o produto tenha significado. Em (c) e (d), Q supõe-se não singular.
 - (a) $(P + Q)' = P' + Q'$.
 - (b) $(PQ)' = PQ' + P'Q$.
 - (c) $(Q^{-1})' = -Q^{-1}Q'Q^{-1}$.
 - (d) $(PQ^{-1})' = -PQ^{-1}Q'Q^{-1} + P'Q^{-1}$.
3. (a) Seja P uma função matricial derivável. Provar que as derivadas de P^2 e P^3 são dadas pelas fórmulas

$$(P^2)' = PP' + P'P, \quad (P^3)' = P^2P' + PP'P + P'P^2.$$

- (b) Escrever uma fórmula geral para a derivação de P^k e demonstrá-la por indução

4. Seja P uma função matricial derivável e g uma função escalar derivável cujo contradomínio é um subconjunto do domínio de P . Definir a função composta $F(t) = P[g(t)]$ e provar a regra de derivação $F'(t) = g'(t)P'[g(t)]$.
5. Demonstrar o teorema: Se $P'(t) = O$ para todo t pertencente a um intervalo aberto (a, b) , então a função matricial P é constante em (a, b) .
6. Estabelecer e demonstrar generalizações dos primeiro e segundo teoremas fundamentais do cálculo para funções matriciais.
7. Estabelecer e demonstrar uma fórmula para a integração por partes na qual os integrandos são funções matriciais.
8. Provar as propriedades seguintes da norma de matrizes

$$\|A + B\| \leq \|A\| + \|B\|, \quad \|cA\| = |c| \|A\|.$$

9. Se uma função matricial P é integrável num intervalo $[a, b]$ provar que

$$\left\| \int_a^b P(t) dt \right\| \leq \int_a^b \|P(t)\| dt.$$

10. Seja D uma matriz diagonal $n \times n$, por exemplo $D = \text{diag}(\lambda_1, \dots, \lambda_n)$. Provar que a série de matrizes $\sum_{k=0}^{\infty} D^k/k!$ converge e é também uma matriz diagonal

$$\sum_{k=0}^{\infty} \frac{D^k}{k!} = \text{diag}(e^{\lambda_1}, \dots, e^{\lambda_n}).$$

(O termo correspondente a $k=0$ subentende-se que é a matriz identidade I .)

11. Seja D uma matriz diagonal $n \times n$, $D = \text{diag}(\lambda_1, \dots, \lambda_n)$. Se a série matricial $\sum_{k=0}^{\infty} c_k D^k$ converge, provar que

$$\sum_{k=0}^{\infty} c_k D^k = \text{diag} \left(\sum_{k=0}^{\infty} c_k \lambda_1^k, \dots, \sum_{k=0}^{\infty} c_k \lambda_n^k \right).$$

12. Supor que a série matricial $\sum_{k=1}^{\infty} C_k$ converge, com cada C_k uma matriz $n \times n$. Provar que a série matricial $\sum_{k=1}^{\infty} (AC_k B)$ também converge e que a sua soma é a matriz

$$A \left(\sum_{k=1}^{\infty} C_k \right) B.$$

Aquí A e B são matrizes tais que os produtos $AC_k B$ tem significado.

7.5. A matriz exponencial

Aplicando o teorema 7.2 é fácil provar que a série matricial

$$(7.12) \quad \sum_{k=0}^{\infty} \frac{A^k}{k!}$$

Converge para qualquer matriz quadrada A de elementos reais ou complexos. (O termo correspondente a $k = 0$ subentende-se ser a matriz identidade I .) A norma de cada termo satisfaz à desigualdade

$$\left\| \frac{A^k}{k!} \right\| \leq \frac{\|A\|^k}{k!}.$$

Uma vez que $\sum a^k/k!$ converge para todo real a , o teorema 7.2 implica que a série em (7.12) converge para qualquer matriz quadrada A .

DEFINIÇÃO DA EXPONENCIAL DE UMA MATRIZ. Para qualquer matriz A , $n \times n$, de elementos reais ou complexos, define-se a exponencial e^A como sendo a matriz $n \times n$ definida pela série convergente (7.12), isto é,

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}.$$

Observe-se que esta definição implica $e^0 = I$, onde O é a matriz nula. Com auxílio das equações diferenciais estudar-se-ão outras propriedades da exponencial de uma matriz.

7.6. A equação diferencial verificada por e^{tA}

Sejam t um número real, A uma matriz $n \times n$ e $E(t)$ a matriz $n \times n$

$$E(t) = e^{tA}.$$

Vamos considerar A fixo e estudar esta matriz como uma função de t . Em primeiro lugar vamos obter uma equação diferencial à qual satisfaça E .

TEOREMA 7.3. Para cada real t a função matricial E definida por $E(t) = e^{tA}$ satisfaz à equação diferencial matricial

$$E'(t) = E(t)A = AE(t).$$

Demonstração. Da definição de exponencial de uma matriz tem-se

$$E(t) = \sum_{k=0}^{\infty} \frac{(tA)^k}{k!} = \sum_{k=0}^{\infty} \frac{t^k A^k}{k!}.$$

Seja $c_{ij}^{(k)}$ o elemento ij de A^k . Então o elemento ij de $t^k A^k/k!$ é $t^k c_{ij}^{(k)}/k!$. Por conseguinte, da definição de uma série matricial, tem-se

$$(7.13) \quad \sum_{k=0}^{\infty} \frac{t^k A^k}{k!} = \left[\sum_{k=0}^{\infty} \frac{t^k}{k!} c_{ij}^{(k)} \right].$$

Cada elemento do segundo membro de (7.13) é uma série de potências em t , convergente para t qualquer. Portanto a sua derivada existe para todo t e é dada pela série derivada.

$$\sum_{k=1}^{\infty} \frac{k t^{k-1}}{k!} c_{ij}^{(k)} = \sum_{k=0}^{\infty} \frac{t^k}{k!} c_{ij}^{(k+1)}.$$

Isto prova que a derivada $E'(t)$ existe e é dada pela série matricial

$$E'(t) = \sum_{k=0}^{\infty} \frac{t^k A^{k+1}}{k!} = \left(\sum_{k=0}^{\infty} \frac{t^k A^k}{k!} \right) A = E(t)A.$$

Na última equação aplicamos a propriedade $A^{k+1} = A^k A$. Porque A comuta com A^k poderíamos também escrever $A^{k+1} = A A^k$ para obter a relação $E'(t) = A E(t)$, o que completa a demonstração.

Nota: A demonstração anterior também prova que A comuta com e^{tA} .

7.7. Teorema da unicidade para a equação diferencial matricial $F'(t) = AF(t)$

A seguir vamos ocupar-nos de um teorema de unicidade que caracteriza todas as soluções da equação diferencial matricial $F'(t) = AF(t)$. A demonstração faz uso do seguinte teorema:

TEOREMA 7.4. NÃO SINGULARIDADE DE e^{tA} . Para qualquer matriz A , $n \times n$, e qualquer escalar t tem-se

$$(7.14) \quad e^{tA} e^{-tA} = I.$$

Por conseguinte e^{tA} é não singular, e a sua inversa é e^{-tA} .

Demonstração. Seja F a função matricial definida para todo real t pela equação

$$F(t) = e^{tA} e^{-tA}.$$

Vamos provar que $F(t)$ é a matriz identidade I , demonstrando que a derivada $F'(t)$ é a matriz nula. Derivando F como um produto, usando o resultado do teorema 7.3, encontramos

$$\begin{aligned} F'(t) &= e^{tA}(e^{-tA})' + (e^{tA})'e^{-tA} = e^{tA}(-Ae^{-tA}) + Ae^{tA}e^{-tA} \\ &= -Ae^{tA}e^{-tA} + Ae^{tA}e^{-tA} = 0, \end{aligned}$$

já que A permuta com e^{tA} . Portanto, segundo o teorema da derivada nula, F é uma matriz constante. Mas $F(0) = e^{0A}e^{-0A} = I$, pelo que $F(t) = I$ para todo t . Isto demonstra (7.14).

TEOREMA 7.5. TEOREMA DE UNICIDADE. *Sejam A e B duas matrizes constantes do tipo $n \times n$. A única função matricial F , $n \times n$, satisfazendo ao problema do valor inicial*

$$F'(t) = AF(t), \quad F(0) = B$$

para $-\infty < t < +\infty$ é

$$(7.15) \quad F(t) = e^{tA}B.$$

Demonstração: Em primeiro lugar observe-se que $e^{tA}B$ é uma solução. Seja F uma solução qualquer e considere-se a função matricial

$$G(t) = e^{-tA}F(t).$$

Derivando este produto obtemos

$$G'(t) = e^{-tA}F'(t) - Ae^{-tA}F(t) = e^{-tA}AF(t) - e^{-tA}AF(t) = 0.$$

Por conseguinte $G(t)$ é uma matriz constante,

$$G(t) = G(0) = F(0) = B.$$

quer dizer $e^{-tA}F(t) = B$. Multiplicando por e^{tA} e aplicando (7.14) obtemos (7.15).

Nota: O mesmo tipo de demonstração mostra que $F(t) = Be^{tA}$ é a única solução do problema do valor inicial

$$F'(t) = F(t)A, \quad F(0) = B.$$

7.8. Regra do produto de exponenciais de matrizes

A regra do produto de exponenciais $e^A e^B = e^{A+B}$ nem sempre é verdadeira para exponenciais de matrizes. Um contra exemplo é dado no Exercício 13 da Secção 7.12. Porém, não é difícil provar que a fórmula é verdadeira para matrizes A e B que sejam permutáveis.

TEOREMA 7.6. *Se A e B são duas matrizes permutáveis, $AB = BA$, então tem-se*

$$(7.16) \quad e^{A+B} = e^A e^B$$

Demonstração: Da igualdade $AB = BA$ deduzimos que

$$A^2B = A(BA) = (AB)A = (BA)A = BA^2,$$

pelo que B permuta com A^2 . Por indução, B permuta com qualquer potência de A . Escrevendo e^{tA} como uma série de potências encontramos que B também permuta com e^{tA} para qualquer real t .

Seja agora F a função matricial definida pela equação

$$F(t) = e^{t(A+B)} - e^{tA}e^{tB}.$$

Derivando $F(t)$ e tendo em conta que B permuta com e^{tA} encontramos

$$\begin{aligned} F'(t) &= (A+B)e^{t(A+B)} - Ae^{tA}e^{tB} - e^{tA}Be^{tB} \\ &= (A+B)e^{t(A+B)} - (A+B)e^{tA}e^{tB} = (A+B)F(t). \end{aligned}$$

Pelo teorema de unicidade temos

$$F(t) = e^{t(A+B)}F(0).$$

Mas $F(0) = O$, pelo que $F(t) = O$ para todo t . Por conseguinte

$$e^{t(A+B)} = e^{tA}e^{tB}.$$

Quando $t = 1$ obtemos (7.16).

EXEMPLO. As matrizes sA e tA permutam, quaisquer que sejam os escalares s e t . Logo temos

$$e^{sA}e^{tA} = e^{(s+t)A}.$$

7.9. Teoremas de existência e unicidade para sistemas lineares homogêneos

A equação diferencial vectorial $Y'(t) = AY(t)$, onde A é uma matriz constante $n \times n$ e Y é uma função vectorial n dimensional (considerada como uma matriz coluna $n \times 1$), diz-se um *sistema linear homogêneo com coeficientes constantes*. Utilizaremos a exponencial de uma matriz para dar uma fórmula explícita para a solução de um tal sistema.

TEOREMA 7.7. *Sejam A uma dada matriz constante, $n \times n$ e B um vector n dimensional dado. O problema de valor inicial*

$$(7.17) \quad Y'(t) = AY(t), \quad Y(0) = B,$$

tem uma só solução no intervalo $-\infty < t < +\infty$. Esta solução é dada pela fórmula

$$(7.18) \quad Y(t) = e^{tA}B.$$

Mais geralmente, a única solução do problema de valor inicial

$$Y'(t) = AY(t), \quad Y(a) = B,$$

$$\text{é } Y(t) = e^{(t-a)A}B.$$

Demonstração. A derivação de (7.18) dá-nos $Y'(t) = Ae^{tA}B = AY(t)$. Uma vez que $Y(0) = B$, esta é uma solução do problema do valor inicial (7.17).

Para provar que é a única solução raciocinamos como na demonstração do teorema 7.5. Seja $Z(t)$ outra função vectorial satisfazendo a $Z'(t) = AZ(t)$ com $Z(0) = B$ e seja $G(t) = e^{-tA}Z(t)$. Então é fácil verificar que $G'(t) = 0$, pelo que $G(t) = G(0) = Z(0) = B$. Por outras palavras, $e^{-tA}Z(t) = B$, pelo que $Z(t) = e^{tA}B = Y(t)$. Do mesmo modo se pode tratar o caso mais geral com o valor inicial $Y(a) = B$.

7.10. O problema do cálculo de e^{tA}

Embora o teorema 7.7 dê uma fórmula explícita para a solução de um sistema homogêneo com coeficientes constantes, ainda permanece o problema do cálculo efectivo de e^{tA} . Se fossemos calcular e^{tA} directamente a partir da definição pela série, teríamos que calcular todas as potências A^k para $k = 0, 1, 2, \dots$, e depois calcular a soma de cada série $\sum_{k=0}^{\infty} t^k c_{ij}^{(k)} / k!$, onde $c_{ij}^{(k)}$ é o elemento ij de A^k . Esta seria, na generalidade, uma tarefa desanimadora, a menos que A seja uma matriz cujas potências possam calcular-se facilmente. Por exemplo, se A é uma matriz diagonal,

$$A = \text{diag}(\lambda_1, \dots, \lambda_n),$$

então cada potência de A é também uma matriz diagonal, pois

$$A^k = \text{diag}(\lambda_1^k, \dots, \lambda_n^k).$$

Por conseguinte, nesta hipótese e^{tA} é uma matriz diagonal definida por

$$e^{tA} = \text{diag} \left(\sum_{k=0}^{\infty} \frac{t^k}{k!} \lambda_1^k, \dots, \sum_{k=0}^{\infty} \frac{t^k}{k!} \lambda_n^k \right) = \text{diag}(e^{t\lambda_1}, \dots, e^{t\lambda_n}).$$

Outro caso fácil de manejar é aquele em que A é uma matriz que pode diagonalizar-se. Por exemplo, se existe uma matriz não singular C tal que $C^{-1}AC$ seja uma matriz diagonal, seja $C^{-1}AC = D$, então temos $A = CDC^{-1}$, donde resulta

$$A^2 = (CDC^{-1})(CDC^{-1}) = CD^2C^{-1},$$

e mais geralmente,

$$A^k = CD^kC^{-1}.$$

Portanto neste caso temos

$$e^{tA} = \sum_{k=0}^{\infty} \frac{t^k A^k}{k!} = \sum_{k=0}^{\infty} \frac{t^k}{k!} CD^kC^{-1} = C \left(\sum_{k=0}^{\infty} \frac{t^k D^k}{k!} \right) C^{-1} = Ce^{tD}C^{-1}.$$

Aqui a dificuldade reside na determinação de C e da sua inversa. Desde que estas sejam conhecidas, e^{tA} calcula-se com facilidade.

Evidentemente que nem toda a matriz é diagonalizável, pelo que a utilidade das observações precedentes é limitada.

EXEMPLO 1. Calcular e^{tA} para a matriz 2×2 , $A = \begin{bmatrix} 5 & 4 \\ 1 & 2 \end{bmatrix}$.

Resolução. Esta matriz tem valores próprios distintos $\lambda_1 = 6$, $\lambda_2 = 1$, pelo que existe uma matriz não singular $C = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ tal que $C^{-1}AC = D$, onde $D = \text{diag}(\lambda_1, \lambda_2) = \begin{bmatrix} 6 & 0 \\ 0 & 1 \end{bmatrix}$. Pela determinação de C podemos escrever $AC = CD$, ou

$$\begin{bmatrix} 5 & 4 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 6 & 0 \\ 0 & 1 \end{bmatrix}.$$

Multiplicando as matrizes, verificamos que esta equação é satisfeita por escalares a , b , c e d quaisquer, desde que $a = 4c$, $b = -d$. Fazendo $c = d = 1$ escolhemos

$$C = \begin{bmatrix} 4 & -1 \\ 1 & 1 \end{bmatrix}, \quad C^{-1} = \frac{1}{5} \begin{bmatrix} 1 & 1 \\ -1 & 4 \end{bmatrix}.$$

Consequentemente

$$\begin{aligned} e^{tA} &= Ce^{tD}C^{-1} = \frac{1}{5} \begin{bmatrix} 4 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} e^{6t} & 0 \\ 0 & e^t \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 4 \end{bmatrix} \\ &= \frac{1}{5} \begin{bmatrix} 4 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} e^{6t} & e^{6t} \\ -e^t & 4e^t \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 4e^{6t} + e^t & 4e^{6t} - 4e^t \\ e^{6t} - e^t & e^{6t} + 4e^t \end{bmatrix}. \end{aligned}$$

EXEMPLO 2. Resolver o sistema linear

$$y_1' = 5y_1 + 4y_2$$

$$y_2' = y_1 + 2y_2$$

sujeito às condições iniciais $y_1(0) = 2$, $y_2(0) = 3$.

Resolução. Na forma matricial o sistema pode escrever-se

$$Y'(t) = AY(t), \quad Y(0) = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \quad \text{onde} \quad A = \begin{bmatrix} 5 & 4 \\ 1 & 2 \end{bmatrix}.$$

Pelo teorema 7.7 a solução é $Y(t) = e^{tA} Y(0)$. Recorrendo à matriz e^{tA} , calculada no Exemplo 1, encontramos

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 4e^{6t} + e^t & 4e^{6t} - 4e^t \\ e^{6t} - e^t & e^{6t} + 4e^t \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

donde se obtém

$$y_1 = 4e^{6t} - 2e^t, \quad y_2 = e^{6t} + 2e^t.$$

Conhecem-se vários métodos para o cálculo de e^{tA} quando A não pode ser diagonalizada. A maior parte desses métodos é bastante mais complicada e exige transformações matriciais prévias, cuja natureza depende das multiplicidades dos valores próprios de A . Numa secção posterior discutiremos um método prático e directo de cálculo de e^{tA} , o qual se pode usar quer A seja ou não diagonalizável. É válido para todas as matrizes A e não exige transformações preliminares de qualquer espécie. Este método foi desenvolvido por E. J. Putzer num artigo publicado no *American Mathematical Monthly*, Vol. 73 (1966), pg. 2-7. Baseia-se o método num famoso teorema atribuído a Arthur Cayley (1821-1895) e William Rowan Hamilton (1805-1865), o qual afirma que toda a matriz quadrada satisfaz à sua equação característica. Em primeiro lugar vamos demonstrar o teorema de Cayley-Hamilton e em seguida aplicá-lo à obtenção da fórmula de Putzer para o cálculo de e^{tA} .

7.11. O teorema de Cayley-Hamilton

TEOREMA 7.8. O TEOREMA DE CAYLEY-HAMILTON. Se A é uma matriz $n \times n$ é

$$(7.19) \quad f(\lambda) = \det(\lambda I - A) = \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0$$

o seu polinómio característico, então $f(A) = O$, isto é, A satisfaz à equação

$$(7.20) \quad A^n + c_{n-1}A^{n-1} + \cdots + c_1A + c_0I = O.$$

Demonstração. A demonstração baseia-se no teorema 3.12 o qual estabelece que para qualquer matriz quadrada A se tem

$$(7.21) \quad A(\text{Cal } A)^t = (\det A)I.$$

Aplicamos esta fórmula com A substituído por $\lambda I - A$. Uma vez que $\det(\lambda I - A) = f(\lambda)$, a equação (7.21) escreve-se

$$(7.22) \quad (\lambda I - A)\{\text{Cal}(\lambda I - A)\}^t = f(\lambda)I.$$

Esta equação é válida para qualquer real λ . A ideia da demonstração consiste em fazer ver que também é válida quando λ se substitui por A .

Os elementos da matriz $\text{Cal}(\lambda I - A)$ são os complementos algébricos de $\lambda I - A$. Excepto para um factor ± 1 , cada tal complemento algébrico é o determinante de uma matriz menor de $\lambda I - A$ de ordem $n - 1$. Deste modo, cada elemento de $\text{Cal}(\lambda I - A)$ e consequentemente de $\{\text{Cal}(\lambda I - A)\}^t$, é um polinómio em λ de grau $\leq n - 1$. Por conseguinte

$$\{\text{Cal}(\lambda I - A)\}^t = \sum_{k=0}^{n-1} \lambda^k B_k,$$

onde cada coeficiente B_k é uma matriz $n \times n$ com elementos escalares. Aplicando isto em (7.22) obtemos a relação

$$(7.23) \quad (\lambda I - A) \sum_{k=0}^{n-1} \lambda^k B_k = f(\lambda)I$$

o qual pode de novo escrever-se na forma

$$(7.24) \quad \lambda^n B_{n-1} + \sum_{k=1}^{n-1} \lambda^k (B_{k-1} - AB_k) - AB_0 = \lambda^n I + \sum_{k=1}^{n-1} \lambda^k c_k I + c_0 I.$$

Igualando agora os coeficientes das potências iguais de λ em (7.24) obtemos as equações

$$(7.25) \quad \begin{aligned} B_{n-1} &= I \\ B_{n-2} - AB_{n-1} &= c_{n-1}I \\ &\vdots \\ B_0 - AB_1 &= c_1I \\ -AB_0 &= c_0I. \end{aligned}$$

Esta operação é possível porque (7.24) é equivalente a n^2 equações escalares, em cada uma das quais podemos igualar os coeficientes de potências semelhantes de λ . Multipliquemos ainda as equações (7.25) sucessivamente por $A^n, A^{n-1}, \dots, A, I$ e adicionemos os resultados. Os termos do primeiro membro anulam-se e obtemos

$$O = A^n + c_{n-1}A^{n-1} + \cdots + c_1A + c_0I.$$

Está pois demonstrado o teorema de Cayley-Hamilton.

Nota: Hamilton demonstrou o teorema em 1853 para uma classe especial de matrizes. Alguns anos mais tarde, Cayley anunciou que o teorema era verdadeiro para qualquer tipo de matrizes, mas não apresentou qualquer demonstração.

EXEMPLO. A matriz $A = \begin{bmatrix} 5 & 4 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 2 \end{bmatrix}$ tem como polinômio característico

$$f(\lambda) = (\lambda - 1)(\lambda - 2)(\lambda - 6) = \lambda^3 - 9\lambda^2 + 20\lambda - 12.$$

O teorema de Cayley-Hamilton estabelece que A satisfaz à equação

$$(7.26) \quad A^3 - 9A^2 + 20A - 12I = O.$$

Esta equação pode usar-se para exprimir A^3 e todas as potências superiores de A em função de I , A e A^2 . Por exemplo, temos

$$A^3 = 9A^2 - 20A + 12I,$$

$$\begin{aligned} A^4 &= 9A^3 - 20A^2 + 12A = 9(9A^2 - 20A + 12I) - 20A^2 + 12A \\ &= 61A^2 - 168A + 108I. \end{aligned}$$

Também se pode utilizar para exprimir A^{-1} como um polinômio em A . De (7.26) resulta também $A(A^2 - 9A + 20I) = 12I$, e obtemos

$$A^{-1} = \frac{1}{12}(A^2 - 9A + 20I).$$

7.12. Exercícios

Em cada um dos Exercícios 1 a 4, (a) exprimir A^{-1} , A^2 e todas as potências superiores de A como uma combinação linear de I e A . (O teorema de Cayley-Hamilton pode auxiliar na resolução); (b) Calcular e^{tA} .

$$1. A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \quad 2. A = \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}, \quad 3. A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad 4. A = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}.$$

$$5. (a) \text{ Se } A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \text{ provar que } e^{tA} = \begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix}.$$

(b) Determinar a fórmula correspondente para e^{tA} quando $A = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}$, a, b reais.

6. Se $F(t) = \begin{bmatrix} t & t-1 \\ 0 & 1 \end{bmatrix}$, provar que $e^{F(t)} = eF(e^{t-1})$.

7. Se $A(t)$ é uma função escalar de t , a derivada de $e^{A(t)}$ é $e^{A(t)}A'(t)$. Calcular a derivada de $e^{A(t)}$ quando $A(t) = \begin{bmatrix} 1 & t \\ 0 & 0 \end{bmatrix}$ e mostrar que o resultado não é igual a nenhum dos dois produtos $e^{A(t)}A'(t)$ ou $A'(t)e^{A(t)}$.

Em cada um dos Exercícios 8, 9, 10, (a) calcular A^n , e exprimir A^3 em função de I, A, A^2 .
(b) Calcular e^{tA} .

$$8. A = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad 9. A = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad 10. A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

$$11. \text{ Se } A = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \text{ exprimir } e^{tA} \text{ como uma combinação linear de } I, A, A^2.$$

$$12. \text{ Se } A = \begin{bmatrix} 0 & 1 & 0 \\ 2 & 0 & 2 \\ 0 & 1 & 0 \end{bmatrix}, \text{ provar que } e^A = \begin{bmatrix} x^2 & xy & y^2 \\ 2xy & x^2 + y^2 & 2xy \\ y^2 & xy & x^2 \end{bmatrix}, \text{ com } x = \text{ch } 1 \text{ e } y = \text{sh } 1.$$

13. Este exemplo mostra que a equação $e^{A+B} = e^A e^B$ nem sempre é verdadeira para as exponenciais de matrizes. Calcular cada uma das matrizes $e^A e^B$, $e^B e^A$, e^{A+B} quando $A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ e $B = \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}$ e observar que os três resultados são distintos.

7.13. Método de Putzer para o cálculo de e^{tA}

O teorema de Cayley-Hamilton mostra que a n -ésima potência de qualquer matriz A , $n \times n$, se pode exprimir como uma combinação linear das potências inferiores $I, A, A^2, \dots, A^{n-1}$. Resulta que cada uma das potências superiores A^{n+1}, A^{n+2}, \dots , pode também exprimir-se como uma combinação de $I, A, A^2, \dots, A^{n-1}$. Portanto, nas séries infinitas definindo e^{tA} , cada termo $t^k A^k / k!$ com $k \geq n$ é uma combinação linear de $t^k I, t^k A, t^k A^2, \dots, t^k A^{n-1}$. Por conseguinte pode pensar-se que e^{tA} possa ser expresso como um polinómio em A da forma

$$(7.27) \quad e^{tA} = \sum_{k=0}^{n-1} q_k(t) A^k,$$

onde os coeficientes escalares $q_k(t)$ dependem de t . Putzer desenvolveu dois métodos úteis para exprimir e^{tA} como um polinómio em A . O teorema que a seguir se indica descreve o mais simples dos dois métodos.

TEOREMA 7.9. *Sejam $\lambda_1, \dots, \lambda_n$ os valores próprios de uma matriz A , $n \times n$ e defina-se uma sucessão de polinómios em A do modo seguinte:*

$$(7.28) \quad P_0(A) = I, \quad P_k(A) = \prod_{m=1}^k (A - \lambda_m I), \quad \text{para } k = 1, 2, \dots, n.$$

Então tem-se

$$(7.29) \quad e^{tA} = \sum_{k=0}^{n-1} r_{k+1}(t) P_k(A),$$

onde os coeficientes escalares $r_1(t), \dots, r_n(t)$ são determinados por recorrência a partir do sistema de equações diferenciais lineares

$$(7.30) \quad \begin{aligned} r'_1(t) &= \lambda_1 r_1(t), & r_1(0) &= 1, \\ r'_{k+1}(t) &= \lambda_{k+1} r_{k+1}(t) + r_k(t), & r_{k+1}(0) &= 0, \end{aligned} \quad (k = 1, 2, \dots, n-1).$$

Nota: A equação (7.29) não exprime e^{tA} directamente em potências de A , como se indica em (7.27), mas como uma combinação linear dos polinómios $P_0(A), P_1(A), \dots, P_{n-1}(A)$. Estes polinómios calculam-se facilmente desde que os valores próprios de A estejam determinados. Também os coeficientes $r_1(t), \dots, r_n(t)$ em (7.30) são de cálculo fácil. Embora isto exija a resolução de um sistema de equações diferenciais lineares, este sistema particular tem uma matriz triangular e as soluções podem determinar-se de forma sucessiva.

Demonstração. Sejam $r_1(t), \dots, r_n(t)$ funções escalares definidas por (7.30) e seja F uma função matricial definida por

$$(7.31) \quad F(t) = \sum_{k=0}^{n-1} r_{k+1}(t) P_k(A).$$

Observe-se que $F(0) = r_1(0) P_0(A) = I$. Provaremos que $F(t) = e^{tA}$, provando que F satisfaz à mesma equação diferencial que e^{tA} , nomeadamente, $F'(t) = AF(t)$.

Derivando (7.31) e utilizando as fórmulas de recorrência (7.30) obtemos

$$F'(t) = \sum_{k=0}^{n-1} r'_{k+1}(t) P_k(A) = \sum_{k=0}^{n-1} \{r_k(t) + \lambda_{k+1} r_{k+1}(t)\} P_k(A),$$

onde $r_0(t)$ é, por definição 0. À igualdade anterior podemos dar a forma

$$F'(t) = \sum_{k=0}^{n-2} r_{k+1}(t)P_{k+1}(A) + \sum_{k=0}^{n-1} \lambda_{k+1} r_{k+1}(t)P_k(A),$$

e substituindo-lhe $\lambda_n F(t) = \sum_{k=0}^{n-1} \lambda_n r_{k+1}(t)P_k(A)$ obtemos a igualdade

$$(7.32) \quad F'(t) - \lambda_n F(t) = \sum_{k=0}^{n-2} r_{k+1}(t) \{P_{k+1}(A) + (\lambda_{k+1} - \lambda_n)P_k(A)\}.$$

Mas de (7.28) vê-se que $P_{k+1}(A) = (A - \lambda_{k+1}I)P_k(A)$, pelo que

$$\begin{aligned} P_{k+1}(A) + (\lambda_{k+1} - \lambda_n)P_k(A) &= (A - \lambda_{k+1}I)P_k(A) + (\lambda_{k+1} - \lambda_n)P_k(A) \\ &= (A - \lambda_n I)P_k(A). \end{aligned}$$

Por conseguinte a equação (7.32) escreve-se

$$\begin{aligned} F'(t) - \lambda_n F(t) &= (A - \lambda_n I) \sum_{k=0}^{n-2} r_{k+1}(t)P_k(A) = (A - \lambda_n I) \{F(t) - r_n(t)P_{n-1}(A)\} \\ &= (A - \lambda_n I)F(t) - r_n(t)P_n(A). \end{aligned}$$

O teorema de Cayley-Hamilton implica que $P_n(A) = O$, pelo que a última equação vem

$$F'(t) - \lambda_n F(t) = (A - \lambda_n I)F(t) = AF(t) - \lambda_n F(t),$$

da qual resulta $F'(t) = AF(t)$. Uma vez que $F(0) = I$, o teorema da unicidade (teorema 7.7) mostra que $F(t) = e^{At}$.

EXEMPLO 1. Exprimir e^{At} como uma combinação linear de I e A , sendo A uma matriz 2×2 com ambos os valores próprios iguais a λ .

Resolução. Escrevendo $\lambda_1 = \lambda_2 = \lambda$, temos que resolver o sistema de equações diferenciais

$$\begin{aligned} r_1'(t) &= \lambda r_1(t), & r_1(0) &= 1, \\ r_2'(t) &= \lambda r_2(t) + r_1(t), & r_2(0) &= 0. \end{aligned}$$

Resolvendo sucessivamente estas equações diferenciais encontramos

$$r_1(t) = e^{\lambda t}, \quad r_2(t) = te^{\lambda t}.$$

Posto que $P_0(A) = I$ e $P_1(A) = A - \lambda I$, a fórmula pedida para e^{tA} é

$$(7.33) \quad e^{tA} = e^{\lambda t}I + te^{\lambda t}(A - \lambda I) = e^{\lambda t}(1 - \lambda t)I + te^{\lambda t}A.$$

EXEMPLO 2. Resolver o Exemplo 1 se os valores próprios de A são λ e μ , com $\lambda \neq \mu$.

Resolução. Neste caso o sistema de equações diferenciais é

$$\begin{aligned} r_1'(t) &= \lambda r_1(t), & r_1(0) &= 1, \\ r_2'(t) &= \mu r_2(t) + r_1(t), & r_2(0) &= 0. \end{aligned}$$

As suas soluções são dadas por

$$r_1(t) = e^{\lambda t}, \quad r_2(t) = \frac{e^{\lambda t} - e^{\mu t}}{\lambda - \mu}.$$

Visto que $P_0(A) = I$ e $P_1(A) = A - \lambda I$, a fórmula pedida para e^{tA} é

$$(7.34) \quad e^{tA} = e^{\lambda t}I + \frac{e^{\lambda t} - e^{\mu t}}{\lambda - \mu}(A - \lambda I) = \frac{\lambda e^{\mu t} - \mu e^{\lambda t}}{\lambda - \mu}I + \frac{e^{\lambda t} - e^{\mu t}}{\lambda - \mu}A.$$

Se os valores próprios λ, μ são números complexos, os exponenciais $e^{\lambda t}$ e $e^{\mu t}$ serão também números complexos. Mas se λ e μ forem complexos conjugados, os escalares multiplicando I e A em (7.34) serão reais. Por exemplo, suponhamos

$$\lambda = \alpha + i\beta, \quad \mu = \alpha - i\beta, \quad \beta \neq 0.$$

Então $\lambda - \mu = 2i\beta$ pelo que a equação (7.34) vem

$$\begin{aligned} e^{tA} &= e^{(\alpha+i\beta)t}I + \frac{e^{(\alpha+i\beta)t} - e^{(\alpha-i\beta)t}}{2i\beta}[A - (\alpha + i\beta)I] \\ &= e^{\alpha t} \left\{ e^{i\beta t}I + \frac{e^{i\beta t} - e^{-i\beta t}}{2i\beta}(A - \alpha I - i\beta I) \right\} \\ &= e^{\alpha t} \left\{ (\cos \beta t + i \operatorname{sen} \beta t)I + \frac{\operatorname{sen} \beta t}{\beta}(A - \alpha I - i\beta I) \right\}. \end{aligned}$$

Os termos contendo i anulam-se e resulta

$$(7.35) \quad e^{tA} = \frac{e^{\alpha t}}{\beta} \{ (\beta \cos \beta t - \alpha \operatorname{sen} \beta t)I + \operatorname{sen} \beta t A \}.$$

7.14. Outros métodos para calcular e^{tA} em casos particulares

O método de Putzer para exprimir e^{tA} como um polinómio em A é completamente geral, devido a ser válido para qualquer matriz quadrada A . Um método geral contudo nem sempre é o mais conveniente para ser usado em determinados casos particulares. Nesta secção vamos apresentar métodos mais simples para o cálculo de e^{tA} , em três casos particulares: (a) Quando todos os valores próprios de A são iguais, (b) quando todos os valores próprios de A são distintos, e (c) quando A tem dois valores próprios distintos e um deles precisamente com multiplicidade 1.

TEOREMA 7.10. *Se A é uma matriz $n \times n$ com todos os valores próprios iguais a λ , então tem-se*

$$(7.36) \quad e^{tA} = e^{\lambda t} \sum_{k=0}^{n-1} \frac{t^k}{k!} (A - \lambda I)^k.$$

Demonstração. Uma vez que as matrizes λI e $t(A - \lambda I)$ permutam, temos

$$e^{tA} = e^{\lambda t I} e^{t(A - \lambda I)} = (e^{\lambda t I}) \sum_{k=0}^{\infty} \frac{t^k}{k!} (A - \lambda I)^k.$$

O teorema de Cayley-Hamilton implica que $(A - \lambda I)^k = O$ para todo $k \geq n$, e o teorema está demonstrado.

TEOREMA 7.11. *Se A é uma matriz $n \times n$, com n valores próprios distintos $\lambda_1, \lambda_2, \dots, \lambda_n$ então tem-se*

$$e^{tA} = \sum_{k=1}^n e^{t\lambda_k} L_k(A),$$

onde $L_k(A)$ é um polinómio em A de grau $n - 1$ dado pela fórmula

$$L_k(A) = \prod_{\substack{j=1 \\ j \neq k}}^n \frac{A - \lambda_j I}{\lambda_k - \lambda_j} \quad \text{para } k = 1, 2, \dots, n.$$

Nota: Os polinómios $L_k(A)$ chamam-se *coeficientes de interpolação de Lagrange*.

Demonstração. Definimos uma função matricial F pela equação

$$(7.37) \quad F(t) = \sum_{k=1}^n e^{t\lambda_k} L_k(A)$$

e verificamos que F satisfaz à equação diferencial $F'(t) = AF(t)$ e à condição inicial $F(0) = I$. De (7.37) vemos que

$$AF(t) - F'(t) = \sum_{k=1}^n e^{t\lambda_k} (A - \lambda_k I) L_k(A).$$

Pelo teorema de Cayley-Hamilton temos $(A - \lambda_k I) L_k(A) = O$ para todo k , pelo que F satisfaz à equação diferencial $F'(t) = AF(t)$.

Para completar a demonstração necessitamos provar que F satisfaz à condição inicial $F(0) = I$, a qual se escreve

$$(7.38) \quad \sum_{k=1}^n L_k(A) = I.$$

No Exercício 16 da Secção 7.15 está esboçada uma demonstração de (7.38).

O teorema que apresentamos a seguir trata do caso em que A tem dois valores próprios distintos, um dos quais tem precisamente multiplicidade 1.

TEOREMA 7.12. *Se A é uma matriz $n \times n$ ($n \geq 3$) com dois valores próprios distintos λ e μ , tendo λ multiplicidade $n - 1$ e μ multiplicidade 1, então verifica-se*

$$e^{tA} = e^{\lambda t} \sum_{k=0}^{n-2} \frac{t^k}{k!} (A - \lambda I)^k + \left\{ \frac{e^{\mu t}}{(\mu - \lambda)^{n-1}} - \frac{e^{\lambda t}}{(\mu - \lambda)^{n-1}} \sum_{k=0}^{n-2} \frac{t^k}{k!} (\mu - \lambda)^k \right\} (A - \lambda I)^{n-1}.$$

Demonstração. Como na demonstração do teorema 7.10 começamos por escrever

$$\begin{aligned} e^{tA} &= e^{\lambda t} \sum_{k=0}^{\infty} \frac{t^k}{k!} (A - \lambda I)^k = e^{\lambda t} \sum_{k=0}^{n-2} \frac{t^k}{k!} (A - \lambda I)^k + e^{\lambda t} \sum_{k=n-1}^{\infty} \frac{t^k}{k!} (A - \lambda I)^k \\ &= e^{\lambda t} \sum_{k=0}^{n-2} \frac{t^k}{k!} (A - \lambda I)^k + e^{\lambda t} \sum_{r=0}^{\infty} \frac{t^{n-1+r}}{(n-1+r)!} (A - \lambda I)^{n-1+r}. \end{aligned}$$

Calculemos agora, na forma finita, a série com índice r , mediante a aplicação do teorema de Cayley-Hamilton. Visto que

$$A - \mu I = A - \lambda I - (\mu - \lambda)I$$

encontramos

$$(A - \lambda I)^{n-1} (A - \mu I) = (A - \lambda I)^n - (\mu - \lambda) (A - \lambda I)^{n-1}.$$

O primeiro membro é O pelo teorema de Cayley-Hamilton, pelo que

$$(A - \lambda I)^n = (\mu - \lambda) (A - \lambda I)^{n-1}.$$

Usando esta relação repetidamente encontramos

$$(A - \lambda I)^{n-1+r} = (\mu - \lambda)^r (A - \lambda I)^{n-1}.$$

Deste modo a série em r vem

$$\begin{aligned}\sum_{r=0}^{\infty} \frac{t^{n-1+r}}{(n-1+r)!} (\mu - \lambda)^r (A - \lambda I)^{n-1} &= \frac{1}{(\mu - \lambda)^{n-1}} \sum_{k=n-1}^{\infty} \frac{t^k}{k!} (\mu - \lambda)^k (A - \lambda I)^{n-1} \\ &= \frac{1}{(\mu - \lambda)^{n-1}} \left\{ e^{t(\mu - \lambda)} - \sum_{k=0}^{n-2} \frac{t^k}{k!} (\mu - \lambda)^k \right\} (A - \lambda I)^{n-1}.\end{aligned}$$

o que completa a demonstração.

A fórmula explícita do teorema 7.12 pode também ser deduzida por aplicação do método de Putzer, mas os pormenores são mais complicados.

As fórmulas explícitas dos teoremas 7.10, 7.11 e 7.12 servem para quaisquer matrizes em que $n \leq 3$. Porque o caso 3×3 aparece muitas vezes na prática, as fórmulas para este caso são apresentadas a seguir para maior comodidade de utilização.

CASO 1. Se uma matriz 3×3 , A , tem os valores próprios $\lambda, \lambda, \lambda$, então

$$e^{tA} = e^{\lambda t} \{ I + t(A - \lambda I) + \frac{1}{2} t^2 (A - \lambda I)^2 \}.$$

CASO 2. Se uma matriz 3×3 , A , tem valores próprios distintos λ, μ, ν , então

$$e^{tA} = e^{\lambda t} \frac{(A - \mu I)(A - \nu I)}{(\lambda - \mu)(\lambda - \nu)} + e^{\mu t} \frac{(A - \lambda I)(A - \nu I)}{(\mu - \lambda)(\mu - \nu)} + e^{\nu t} \frac{(A - \lambda I)(A - \mu I)}{(\nu - \lambda)(\nu - \mu)}.$$

CASO 3. Se uma matriz 3×3 , A , tem valores próprios λ, λ, μ , com $\lambda \neq \mu$, então

$$e^{tA} = e^{\lambda t} \{ I + t(A - \lambda I) \} + \frac{e^{\mu t} - e^{\lambda t}}{(\mu - \lambda)^2} (A - \lambda I)^2 - \frac{te^{\lambda t}}{\mu - \lambda} (A - \lambda I)^2.$$

EXEMPLO. Calcular e^{tA} quando $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & -5 & 4 \end{bmatrix}$.

Resolução: Os valores próprios de A são 1, 1, 2 e portanto a fórmula do Caso 3 dá-nos

$$(7.39) \quad e^{tA} = e^t \{ I + t(A - I) \} + (e^{2t} - e^t)(A - I)^2 - te^t(A - I)^2.$$

Ordenando segundo as potências de A obtemos

$$(7.40) \quad e^{tA} = (-2te^t + e^{2t})I + \{(3t + 2)e^t - 2e^{2t}\}A - \{(t + 1)e^t - e^{2t}\}A^2.$$

Chegados a este ponto podemos calcular $(A - I)^2$ ou A^2 e efectuar as operações indicadas em (7.39) ou (7.40) para escrevermos o resultado como uma matriz 3×3 ,

$$e^{tA} = \begin{bmatrix} -2te^t + e^{2t} & (3t+2)e^t - 2e^{2t} & -(t+1)e^t + e^{2t} \\ -2(t+1)e^t + 2e^{2t} & (3t+5)e^t - 4e^{2t} & -(t+2)e^t + 2e^{2t} \\ -2(t+2)e^t + 4e^{2t} & (3t+8)e^t - 8e^{2t} & -(t+4)e^t + 4e^{2t} \end{bmatrix}.$$

7.15. Exercícios

Para cada uma das matrizes dos Exercícios 1 a 6, exprimir e^{tA} como um polinómio em A .

$$1. A = \begin{bmatrix} 5 & -2 \\ 4 & -1 \end{bmatrix}, \quad 2. A = \begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix}, \quad 3. A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix}.$$

$$4. A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -6 & -11 & -6 \end{bmatrix}, \quad 5. A = \begin{bmatrix} 3 & -1 & 1 \\ 2 & 0 & 1 \\ 1 & -1 & 2 \end{bmatrix}, \quad 6. A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}.$$

7. (a) Uma matriz A , 3×3 , tem todos os seus valores próprios iguais a λ . Demonstrar que

$$e^{tA} = \frac{1}{2}e^{\lambda t} \{ (\lambda^2 t^2 - 2\lambda t + 2)I + (-2\lambda t^2 + 2t)A + t^2 A^2 \}.$$

(b) Determinar a fórmula correspondente se A for uma matriz 4×4 com todos os valores próprios iguais a λ .

Em cada um dos Exercícios 8 a 15, resolver o sistema $Y' = AY$, sujeito à condição inicial dada

$$8. A = \begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix}, \quad Y(0) = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}, \quad 9. A = \begin{bmatrix} -5 & 3 \\ -15 & 7 \end{bmatrix}, \quad Y(0) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

$$10. A = \begin{bmatrix} 3 & -1 & 1 \\ 2 & 0 & 1 \\ 1 & -1 & 2 \end{bmatrix}, \quad Y(0) = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \quad 11. A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \quad Y(0) = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}.$$

$$12. A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -6 & -11 & -6 \end{bmatrix}, \quad Y(0) = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad 13. A = \begin{bmatrix} -2 & 2 & -3 \\ 2 & 1 & -6 \\ -1 & -2 & 0 \end{bmatrix}, \quad Y(0) = \begin{bmatrix} 8 \\ 0 \\ 0 \end{bmatrix}.$$

$$14. A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}, \quad Y(0) = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \quad 15. A = \begin{bmatrix} 0 & 0 & 2 & 0 \\ 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 4 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad Y(0) = \begin{bmatrix} 1 \\ 0 \\ 2 \\ 1 \end{bmatrix}.$$

16. Este exercício esboça uma demonstração da Equação (7.38) utilizada na demonstração do teorema 7.11. Seja $L_k(\lambda)$ um polinómio em λ de grau $n - 1$ definido pela equação

$$L_k(\lambda) = \prod_{\substack{j=1 \\ j \neq k}}^n \frac{\lambda - \lambda_j}{\lambda_k - \lambda_j},$$

onde $\lambda_1, \dots, \lambda_n$ são n escalares distintos.

(a) Provar que

$$L_k(\lambda_i) = \begin{cases} 0 & \text{if } \lambda_i \neq \lambda_k, \\ 1 & \text{if } \lambda_i = \lambda_k. \end{cases}$$

(b) Sejam y_1, \dots, y_n , n escalares arbitrários, e seja

$$p(\lambda) = \sum_{k=1}^n y_k L_k(\lambda).$$

Provar que $p(\lambda)$ é o único polinómio de grau $\leq n - 1$ que satisfaz às n equações

$$p(\lambda_k) = y_k \quad \text{para } k = 1, 2, \dots, n.$$

(c) Provar que $\sum_{k=1}^n L_k(\lambda) = 1$ para todo λ , e verificar que para toda a matriz quadrada A se tem

$$\sum_{k=1}^n L_k(A) = I,$$

onde I é a matriz identidade.

7.16. Sistemas lineares não homogêneas com coeficientes constantes

Consideremos em seguida o problema de valor inicial não homogêneo

$$(7.41) \quad Y'(t) = AY(t) + Q(t), \quad Y(a) = B,$$

um intervalo J . Aqui A é uma matriz constante $n \times n$, Q é uma função vectorial n -dimensional (considerada como uma matriz coluna $n \times 1$) contínua em J , e a é um ponto dado de J . Podemos obter uma fórmula explícita para a solução deste problema pelo mesmo processo usado para tratar o caso escalar.

Multiplicamos em primeiro lugar ambos os membros de (7.41) pela matriz exponencial e^{-tA} e escrevemos de novo a equação diferencial na forma

$$(7.42) \quad e^{-tA}\{Y'(t) - AY(t)\} = e^{-tA}Q(t).$$

O primeiro membro de (7.42) é a derivada do produto $e^{-tA}Y(t)$. Portanto, se integramos ambos os membros de (7.42) de a a x , com $x \in J$, obtemos

$$e^{-xA}Y(x) - e^{-aA}Y(a) = \int_a^x e^{-tA}Q(t) dt.$$

Multiplicando por e^{xA} obtemos a fórmula explícita (7.43) que aparece no teorema seguinte.

TEOREMA 7.13. *Se A é uma matriz constante $n \times n$ e Q uma função vectorial n dimensional contínua em J , então o problema do valor inicial*

$$Y'(t) = AY(t) + Q(t), \quad Y(a) = B,$$

tem uma solução única em J dada pela fórmula

$$(7.43) \quad Y(x) = e^{(x-a)A}B + e^{xA} \int_a^x e^{-tA}Q(t) dt.$$

Como no caso homogêneo, a dificuldade de aplicação desta fórmula na prática reside no cálculo de exponenciais de matrizes.

Observe-se que o primeiro termo, $e^{(x-a)A}B$, é a solução do problema homogêneo $Y'(t) = AY(t)$, $Y(a) = B$. O segundo termo é a solução do problema não homogêneo

$$Y'(t) = AY(t) + Q(t), \quad Y(a) = O.$$

Ilustramos o teorema 7.13 com um exemplo.

EXEMPLO 1. Resolver o problema de valor inicial

$$Y'(t) = AY(t) + Q(t), \quad Y(0) = B,$$

no intervalo $(-\infty, +\infty)$, em que

$$A = \begin{bmatrix} 2 & -1 & 1 \\ 0 & 3 & -1 \\ 2 & 1 & 3 \end{bmatrix}, \quad Q(t) = \begin{bmatrix} e^{2t} \\ 0 \\ te^{2t} \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Resolução. De acordo com o teorema 7.13, a solução é dada por

$$(7.44) \quad Y(x) = e^{xA} \int_0^x e^{-tA}Q(t) dt = \int_0^x e^{(x-t)A}Q(t) dt.$$

Os valores próprios de A são 2, 2 e 4. Para calcular e^{xA} usamos a fórmula do caso III, Secção 7.14, para obtermos

$$\begin{aligned} e^{xA} &= e^{2x}\{I + x(A - 2I)\} + \frac{1}{2}(e^{4x} - e^{2x})(A - 2I)^2 - \frac{1}{2}xe^{2x}(A - 2I)^2 \\ &= e^{2x}\{I + x(A - 2I) + \frac{1}{4}(e^{2x} - 2x - 1)(A - 2I)^2\}. \end{aligned}$$

Podemos substituir x por $x - t$ nesta fórmula para obtermos $e^{(x-t)A}$. Deste modo o integrando em (7.44) é

$$\begin{aligned} e^{(x-t)A}Q(t) &= e^{2(x-t)}\{I + (x-t)(A - 2I) + \frac{1}{4}[e^{2(x-t)} - 2(x-t) - 1](A - 2I)^2\}Q(t) \\ &= e^{2x}\begin{bmatrix} 1 \\ 0 \\ t \end{bmatrix} + (A - 2I)e^{2x}\begin{bmatrix} x - t \\ 0 \\ t(x - t) \end{bmatrix} + \frac{1}{4}(A - 2I)^2e^{2x}\begin{bmatrix} e^{2x}e^{-2t} - 2(x-t) - 1 \\ 0 \\ e^{2x}te^{-2t} - 2t(x-t) - t \end{bmatrix}. \end{aligned}$$

Integrando, vem

$$\begin{aligned} Y(x) = \int_0^x e^{(x-t)A}Q(t) dt &= e^{2x}\begin{bmatrix} x \\ 0 \\ \frac{1}{2}x^2 \end{bmatrix} + (A - 2I)e^{2x}\begin{bmatrix} \frac{1}{2}x^2 \\ 0 \\ \frac{1}{6}x^3 \end{bmatrix} \\ &\quad + \frac{1}{4}(A - 2I)^2e^{2x}\begin{bmatrix} \frac{1}{2}e^{2x} - \frac{1}{2} - x - x^2 \\ 0 \\ \frac{1}{4}e^{2x} - \frac{1}{4} - \frac{1}{2}x - \frac{1}{2}x^2 - \frac{1}{3}x^3 \end{bmatrix}. \end{aligned}$$

Uma vez que se tem

$$A - 2I = \begin{bmatrix} 0 & -1 & 1 \\ 0 & 1 & -1 \\ 2 & 1 & 1 \end{bmatrix} \quad \text{e} \quad (A - 2I)^2 = \begin{bmatrix} 2 & 0 & 2 \\ -2 & 0 & -2 \\ 2 & 0 & 2 \end{bmatrix},$$

encontramos

$$\begin{aligned} Y(x) &= e^{2x}\begin{bmatrix} x \\ 0 \\ \frac{1}{2}x^2 \end{bmatrix} + e^{2x}\begin{bmatrix} \frac{1}{6}x^3 \\ -\frac{1}{6}x^3 \\ x^2 + \frac{1}{6}x^3 \end{bmatrix} + e^{2x}\begin{bmatrix} \frac{3}{8}e^{2x} - \frac{3}{8} - \frac{3}{4}x - \frac{3}{4}x^2 - \frac{1}{6}x^3 \\ -\frac{3}{8}e^{2x} + \frac{3}{8} + \frac{3}{4}x + \frac{3}{4}x^2 + \frac{1}{6}x^3 \\ \frac{3}{8}e^{2x} - \frac{3}{8} - \frac{3}{4}x - \frac{3}{4}x^2 - \frac{1}{6}x^3 \end{bmatrix} \\ &= e^{2x}\begin{bmatrix} \frac{3}{8}e^{2x} - \frac{3}{8} + \frac{1}{4}x - \frac{3}{4}x^2 \\ -\frac{3}{8}e^{2x} + \frac{3}{8} + \frac{3}{4}x + \frac{3}{4}x^2 \\ \frac{3}{8}e^{2x} - \frac{3}{8} - \frac{3}{4}x + \frac{3}{4}x^2 \end{bmatrix}. \end{aligned}$$

As linhas desta matriz são as funções pedidas y_1, y_2, y_3 .

7.17. Exercícios

1. Seja Z uma solução do sistema não homogêneo

$$Z'(t) = AZ(t) + Q(t),$$

num intervalo J , com valor inicial $Z(a)$. Provar que existe unicamente uma solução do sistema não homogêneo

$$Y'(t) = AY(t) + Q(t)$$

em J , com valor inicial $Y(a)$, a qual é dada pela fórmula

$$Y(t) = Z(t) + e^{(t-a)A}\{Y(a) - Z(a)\}.$$

Frequentemente dispõe-se de outros métodos para determinar uma solução particular $Z(t)$ que se assemelha à função dada $Q(t)$. Os exercícios 2, 3, 5 e 7 indicam tais métodos para $Q(t) = C$, $Q(t) = e^{at}C$, $Q(t) = t^m C$, e $Q(t) = (\cos at)C + (\sin at)D$, com C e D vectores constantes. Se a solução particular $Z(t)$ assim obtida não tem o valor inicial pedido, modificamos $Z(t)$ como indicámos no Exercício 1 para obter outra solução $Y(t)$ com o valor inicial pretendido.

2. (a) Sejam A uma matriz constante $n \times n$, B e C vectores constantes n -dimensionais. Provar que a solução do sistema

$$Y'(t) = AY(t) + C, \quad Y(a) = B,$$

em $(-\infty, +\infty)$ é dada pela fórmula

$$Y(x) = e^{(x-a)A}B + \left(\int_0^{x-a} e^{uA} du \right) C.$$

- (b) Se A é não singular, mostrar que o integral da alínea (a) tem o valor $(e^{(x-a)A} - I)A^{-1}C$.
(c) Calcular $Y(x)$ quando

$$A = \begin{bmatrix} -1 & 2 \\ -2 & 3 \end{bmatrix}, \quad C = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad B = \begin{bmatrix} b \\ c \end{bmatrix}, \quad a = 0.$$

3. Sejam A uma matriz constante $n \times n$, B e C vectores constantes n -dimensionais e α um escalar dado.
(a) Provar que o sistema não homogêneo $Z'(t) = AZ(t) + e^{at}C$, tem uma solução $Z(t) = e^{at}B$ se, e só se, $(\alpha I - A)B = C$.
(b) Se α não é um valor próprio de A , provar que o vector B pode sempre escolher-se de maneira que o sistema em (a) tenha uma solução da forma $Z(t) = e^{at}B$.
(c) Se α não é um valor próprio de A , provar que toda a solução do sistema $Y'(t) = AY(t) + e^{at}C$ tem a forma $Y(t) = e^{tA}(Y(0) - B) + e^{at}B$, onde $B = (\alpha I - A)^{-1}C$.
4. Usar o método sugerido pelo Exercício 3 para determinar uma solução do sistema não homogêneo $Y'(t) = AY(t) + e^{at}C$; com

$$A = \begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix}, \quad C = \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \quad Y(0) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

5. Sejam A uma matriz constante $n \times n$ e B e C vectores constantes n -dimensionais e m um inteiro positivo.

(a) Provar que o sistema não homogêneo $Y'(t) = AY(t) + t^m C$, $Y(0) = B$, tem uma solução particular da forma

$$Y(t) = B_0 + tB_1 + t^2B_2 + \cdots + t^mB_m,$$

onde B_0, B_1, \dots, B_m são vectores constantes, se e só se

$$C = -\frac{1}{m!} A^{m+1}B.$$

Determinar os coeficientes B_0, B_1, \dots, B_m para uma tal solução.

(b) Se A é não singular, provar que o vector inicial B pode sempre escolher-se de maneira que o sistema dado em (a) tenha uma solução da forma indicada.

6. Considere o sistema não homogêneo

$$y'_1 = 3y_1 + y_2 + t^3$$

$$y'_2 = 2y_1 + 2y_2 + t^3.$$

(a) Determinar uma solução particular da forma $Y(t) = B_0 + tB_1 + t^2B_2 + t^3B_3$.

(b) Determinar uma solução do sistema com $y_1(0) = y_2(0) = 1$.

7. Sejam A uma matriz constante $n \times n$, B, C, D vectores constantes n -dimensionais e α um número real dado diferente de zero. Provar que o sistema não homogêneo

$$Y'(t) = AY(t) + (\cos \alpha t)C + (\sin \alpha t)D, \quad Y(0) = B,$$

tem uma solução particular da forma

$$Y(t) = (\cos \alpha t)E + (\sin \alpha t)F,$$

com E e F vectores constantes, se e só se

$$(A^2 + \alpha^2 I)B = -(AC + \alpha D).$$

Determinar E e F em função de A, B, C para uma tal solução. Observar que se $A^2 + \alpha^2 I$ é não singular, o vector inicial B pode sempre ser escolhido de tal maneira que o sistema tem uma solução da forma indicada.

8. (a) Determinar uma solução particular do sistema não homogêneo

$$y'_1 = y_1 + 3y_2 + 4 \sin 2t$$

$$y'_2 = y_1 - y_2.$$

(b) Determinar uma solução do sistema com $y_1(0) = y_2(0) = 1$.

Em cada um dos Exercícios 9 a 12, resolver o sistema não homogêneo $Y'(t) = AY(t) + Q(t)$ subordinado à condição inicial dada

$$9. A = \begin{bmatrix} 4 & 1 \\ -2 & 1 \end{bmatrix}, \quad Q(t) = \begin{bmatrix} 0 \\ -2e^t \end{bmatrix}, \quad Y(0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

$$10. A = \begin{bmatrix} -5 & -1 \\ 1 & -3 \end{bmatrix}, \quad Q(t) = \begin{bmatrix} e \\ e^{2t} \end{bmatrix}, \quad Y(0) = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}.$$

$$11. A = \begin{bmatrix} -5 & -1 \\ 2 & -3 \end{bmatrix}, \quad Q(t) = \begin{bmatrix} 7e^t - 27 \\ -3e^t + 12 \end{bmatrix}, \quad Y(0) = \begin{bmatrix} -\frac{1007}{442} \\ \frac{707}{221} \end{bmatrix}.$$

$$12. A = \begin{bmatrix} -1 & -1 & 0 \\ 0 & -1 & -1 \\ 0 & 0 & -1 \end{bmatrix}, \quad Q(t) = \begin{bmatrix} t^2 \\ 2t \\ t \end{bmatrix}, \quad Y(0) = \begin{bmatrix} 6 \\ -2 \\ 1 \end{bmatrix}.$$

7.18. O sistema linear geral $Y'(t) = P(t)Y(t) + Q(t)$

O teorema 7.13 dá-nos uma forma explícita para a solução do sistema linear

$$Y'(t) = AY(t) + Q(t), \quad Y(a) = B,$$

onde A é uma matriz constante $n \times n$ e $Q(t)$, $Y(t)$ são matrizes coluna $n \times 1$. Voltamos agora ao caso mais geral

$$(7.45) \quad Y'(t) = P(t)Y(t) + Q(t); \quad Y(a) = B,$$

onde a matriz $P(t)$, $n \times n$, não é necessariamente constante.

Se P e Q são contínuas num intervalo aberto J , um teorema geral de existência e unicidade que demonstraremos mais adiante diz-nos que para cada a em J e cada vector inicial B existe precisamente uma solução para o problema do valor inicial (7.45). Nesta secção utilizamos este resultado a fim de obtermos uma fórmula para a solução, generalizando o teorema 7.13.

No caso escalar ($n=1$) a equação diferencial (7.45) pode resolver-se do modo seguinte: Seja $A(x) = \int_a^x P(t)dt$ e multipliquemos ambos os membros de (7.45) por $e^{-A(t)}$, escrevendo a equação diferencial na forma

$$(7.46) \quad e^{-A(t)}\{Y'(t) - P(t)Y(t)\} = e^{-A(t)}Q(t).$$

O primeiro membro é a derivada do produto $e^{-A(t)}Y(t)$. Consequentemente podemos integrar os membros de a a x , em que a e x são pontos de J , para obtermos

$$e^{-A(x)}Y(x) - e^{-A(a)}Y(a) = \int_a^x e^{-A(t)}Q(t) dt.$$

Multiplicando por $e^{A(x)}$ obtemos a fórmula

$$(7.47) \quad Y(x) = e^{A(x)} e^{-A(a)} Y(a) + e^{A(x)} \int_a^x e^{-A(t)} Q(t) dt.$$

A única parte deste raciocínio que não se aplica de imediato a funções matriciais é a afirmação de que o primeiro membro de (7.46) é a derivada do produto $e^{-A(t)} Y(t)$. Neste ponto servimo-nos do facto de que a derivada de $e^{-A(t)}$ é $-P(t)e^{-A(t)}$. No caso escalar isto é uma consequência da fórmula de derivação de funções exponenciais:

$$\text{Se } E(t) = e^{A(t)}, \quad \text{então } E'(t) = A'(t)e^{A(t)}.$$

Infelizmente esta fórmula de derivação nem sempre é verdadeira quando A é uma função matricial. Por exemplo, é falsa para a função matricial 2×2 , $A(t) = \begin{bmatrix} 1 & t \\ 0 & 0 \end{bmatrix}$. (Ver

Exercício 7 da Secção 7.12). Por conseguinte torna-se necessário modificar o raciocínio para generalizar a equação (7.47) ao caso matricial.

Admitamos que multiplicamos cada membro de (7.45) por uma matriz $n \times n$, $F(t)$, não especificada. Resulta a relação

$$F(t)Y'(t) = F(t)P(t)Y(t) + F(t)Q(t).$$

Adicionemos agora $F'(t)Y(t)$ a ambos os membros a fim de transformarmos o primeiro membro na derivada do produto $F(t)Y(t)$. Feito isto a última equação dá-nos

$$\{F(t)Y(t)\}' = \{F'(t) + F(t)P(t)\}Y(t) + F(t)Q(t).$$

Se podemos escolher a matriz $F(t)$ de tal maneira que a soma $\{F'(t) + F(t)P(t)\}$ no segundo membro seja a matriz nula, a última equação simplifica-se para

$$\{F(t)Y(t)\}' = F(t)Q(t).$$

Integrando de a a x obtemos

$$F(x)Y(x) - F(a)Y(a) = \int_a^x F(t)Q(t) dt.$$

Se, além disso, a matriz $F(x)$ é não singular, obtemos a fórmula explícita

$$(7.48) \quad Y(x) = F(x)^{-1}F(a)Y(a) + F(x)^{-1} \int_a^x F(t)Q(t) dt.$$

Esta é uma generalização da fórmula escalar (7.47). O processo funcionará se pudermos determinar uma função matricial $F(t)$, $n \times n$, que satisfaça à equação diferencial matricial

$$F'(t) = -F(t)P(t)$$

e seja não singular.

Observe-se que esta equação diferencial é muito semelhante à equação diferencial original (7.45) com $Q(t) = 0$, excepto que a função desconhecida $F(t)$ é uma matriz quadrada em vez de uma matriz coluna. Ainda a função incógnita está multiplicada à direita por $-P(t)$ em vez de o estar à esquerda por $P(t)$.

Provaremos a seguir que a equação diferencial para F tem sempre uma solução não singular. A demonstração dependerá do seguinte teorema para sistemas lineares homogêneos.

TEOREMA 7.14. *Se $A(t)$ é uma função matricial $n \times n$, contínua num intervalo aberto J e se $a \in J$ e B é um vector n -dimensional dado, o sistema linear homogêneo*

$$Y'(t) = A(t)Y(t), \quad Y(a) = B,$$

admite em J uma solução que é um vector n -dimensional.

A demonstração de 7.14 é dada na Secção 7.21. Com auxílio deste teorema podemos demonstrar o seguinte.

TEOREMA 7.15. *Dada uma função matricial P , $n \times n$, contínua um intervalo aberto J , e dado um ponto qualquer a em J , existe uma função matricial F , $n \times n$, que satisfaz à equação diferencial matricial*

$$(7.49) \quad F'(x) = -F(x)P(x)$$

em J com o valor inicial $F(a) = I$. Acresce ainda que $F(x)$ é não singular para todo o x de J .

Demonstração. Seja $Y_k(x)$ uma solução vectorial da equação diferencial

$$Y'_k(x) = -P(x)^t Y_k(x)$$

em J com vector inicial $Y_k(a) = I_k$, onde I_k é a coluna de ordem k da matriz identidade I , $n \times n$. Aqui $P(x)^t$ representa a transposta de $P(x)$. Seja $G(x)$ a matriz $n \times n$ cuja coluna de ordem k é $Y_k(x)$. Então G satisfaz à equação diferencial matricial.

$$(7.50) \quad G'(x) = -P(x)^t G(x)$$

em J , com valor inicial $G(a) = I$. Tomemos agora a transposta de cada membro de (7.50). Uma vez que a transposta de um produto é o produto das transpostas por ordem inversa, obtemos

$$\{G'(x)\}^t = -G(x)^t P(x).$$

Também a transposta da derivada G' é a derivada da transposta G^t . Portanto a matriz $F(x) = G(x)^t$ satisfaz à equação diferencial (7.49) com valor inicial $F(a) = I$.

Demonstremos agora que $F(x)$ é não singular construindo a sua inversa. Seja H uma função matricial $n \times n$, cuja coluna de ordem k é a solução da equação diferencial

$$Y'(x) = P(x)Y(x)$$

com valor inicial $Y(a) = I_k$, a coluna de ordem k de I . Então H satisfaz ao problema de valor inicial

$$H'(x) = P(x)H(x), \quad H(a) = I,$$

em J . O produto $F(x)H(x)$ tem a derivada

$$F(x)H'(x) + F'(x)H(x) = F(x)P(x)H(x) - F(x)P(x)H(x) = 0$$

para todo x em J . Deste modo o produto $F(x)H(x)$ é constante. $F(x)H(x) = F(a)H(a) = I$, pelo que $H(x)$ é a inversa de $F(x)$. Completamos assim a demonstração do teorema.

Os resultados desta secção podem resumir-se no teorema seguinte:

TEOREMA 7.16. *Dada uma função matricial $n \times n$, P , e uma função vectorial n -dimensional Q , ambas contínuas no intervalo aberto J , a solução do problema de valor inicial*

$$(7.51) \quad Y'(x) = P(x)Y(x) + Q(x), \quad Y(a) = B,$$

em J é dada pela fórmula

$$(7.52) \quad Y(x) = F(x)^{-1}Y(a) + F(x)^{-1} \int_a^x F(t)Q(t) dt.$$

A matriz $F(x)$, $n \times n$, é a transposta da matriz cuja coluna de ordem k é a solução do problema de valor inicial

$$(7.53) \quad Y'(x) = -P(x)^t Y(x), \quad Y(a) = I_k,$$

onde I_k é a coluna de ordem k da matriz identidade I .

Embora o teorema 7.16 proporcione uma fórmula explícita para a solução do sistema linear geral (7.51), a fórmula nem sempre é útil para o cálculo da solução devido à dificuldade criada pela determinação da matriz funcional F . A determinação de F exige a solução de n sistemas lineares homogêneos (7.53). Na secção seguinte vamos analisar um método por séries de potências que algumas vezes é usado de preferência na resolução de sistemas lineares homogêneos.

Lembramos ao leitor, uma vez mais, que a demonstração do teorema 7.16 se baseia na do teorema 7.14, o teorema de existência para sistemas lineares homogêneos, o qual ainda não foi demonstrado.

7.19. Resolução de sistemas lineares homogêneos por intermédio de séries de potências.

Consideremos um sistema linear homogêneo

$$(7.54) \quad Y'(x) = A(x)Y(x), \quad Y(0) = B,$$

no qual a matriz dada $A(x)$, $n \times n$, admite um desenvolvimento em série de potências de x convergente em algum intervalo aberto contendo a origem, seja

$$A(x) = A_0 + xA_1 + x^2A_2 + \cdots + x^kA_k + \cdots, \quad \text{para } |x| < r_1,$$

onde os coeficientes A_0, A_1, A_2, \dots , são matrizes $n \times n$ dadas. Tentemos encontrar uma solução em forma de série de potências

$$Y(x) = B_0 + xB_1 + x^2B_2 + \cdots + x^kB_k + \cdots,$$

com os coeficientes vectoriais B_0, B_1, B_2, \dots . Uma vez que $Y(0) = B_0$, a condição inicial será satisfeita tomando $B_0 = B$, o vector inicial prescrito. Para determinar os restantes coeficientes substituímos a série de potências para $Y(x)$ na equação diferencial e igualamos os coeficientes dos termos de igual grau, obtendo o seguinte sistema de equações:

$$(7.55) \quad B_1 = A_0B, \quad (k+1)B_{k+1} = \sum_{r=0}^k A_r B_{k-r} \quad \text{para } k = 1, 2, \dots$$

Estas equações podem ser sucessivamente resolvidas relativamente aos vectores B_1, B_2, \dots . Se a série de potências resultante para $Y(x)$ converge em algum intervalo $|x| < r_2$, então $Y(x)$ será uma solução do problema de valor inicial 7.54 no intervalo $|x| < r$, onde $r = \min\{r_1, r_2\}$.

Por exemplo, se $A(x)$ é uma matriz constante A , então $A_0 = A$ e $A_k = 0$ para $k \geq 1$, pelo que o sistema de equações (7.55) se escreve

$$B_1 = AB, \quad (k+1)B_{k+1} = AB_k \quad \text{para } k \geq 1.$$

Resolvendo estas equações sucessivamente encontramos

$$B_k = \frac{1}{k!} A^k B \quad \text{para } k \geq 1.$$

Consequentemente a série solução neste caso escreve-se

$$Y(x) = B + \sum_{k=1}^{\infty} \frac{x^k}{k!} A^k B = e^{xA} B.$$

Isto está de acordo com o resultado anteriormente obtido para sistemas com coeficientes constantes.

7.20. Exercícios

1. Seja p uma função real e Q uma função matricial $n \times 1$, ambas contínuas num intervalo J . Seja A uma matriz constante $n \times n$. Provar que o problema de valor inicial

$$Y'(x) = p(x)AY(x) + Q(x), \quad Y(a) = B,$$

tem a solução

$$Y(x) = e^{q(x)A} B + e^{q(x)A} \int_a^x e^{-q(t)A} Q(t) dt$$

em J , onde $q(x) = \int_a^x p(t) dt$.

2. Considerar o caso especial do Exercício 1 no qual A é não singular, $a = 0$, $p(x) = 2x$, e $Q(x) = xC$, com C um vector constante. Mostrar que a solução se pode escrever

$$Y(x) = e^{x^2 A} (B + \frac{1}{2} A^{-1} C) - \frac{1}{2} A^{-1} C.$$

3. Sejam $A(t)$ uma função matricial $n \times n$ e $E(t) = e^{At}$. Sejam $Q(t)$, $Y(t)$ e B matrizes coluna $n \times 1$. Supôr que

$$E'(t) = A'(t)E(t)$$

num intervalo aberto J . Se $a \in J$ e se A' e Q são contínuas em J , provar que o problema de valor inicial

$$Y'(t) = A'(t)Y(t) + Q(t), \quad Y(a) = B,$$

tem a seguinte solução em J :

$$Y(x) = e^{A(x)} e^{-A(a)} B + e^{A(x)} \int_a^x e^{-A(t)} Q(t) dt.$$

4. Seja $E(t) = e^{At}$. Este exercício mostra exemplos de funções matriciais $A(t)$ para os quais $E'(t) = A'(t)E(t)$.
- (a) Seja $A(t) = t^r A$, onde A é uma matriz constante $n \times n$ e r um inteiro positivo. Provar que $E'(t) = A'(t)E(t)$ em $(-\infty, +\infty)$.
- (b) Seja $A(t)$ um polinómio em t com coeficientes matriciais, seja

$$A(t) = \sum_{r=0}^m t^r A_r,$$

sendo os coeficientes permutáveis $A_r A_s = A_s A_r$ para quaisquer r e s . Provar que $E'(t) = A'(t)E(t)$ em $(-\infty, +\infty)$.

- (c) Resolver o sistema linear homogéneo

$$Y'(t) = (I + tA)Y(t), \quad Y(0) = B$$

no intervalo $(-\infty, +\infty)$, em que A é uma matriz constante $n \times n$.

5. Suponhamos que a função matricial $n \times n$, $A(x)$, admite um desenvolvimento em série de potências convergente para $|x| < r$. Desenvolver um processo para resolver, mediante um desenvolvimento em série, o seguinte sistema linear homogêneo de segunda ordem:

$$Y''(x) = A(x)Y(x), \quad \text{com} \quad Y(0) = B, \quad Y'(0) = C.$$

6. Consideremos o sistema de segunda ordem $Y''(x) + AY(x) = 0$, com $Y(0) = B$, $Y'(0) = C$, onde A é uma matriz constante $n \times n$. Provar que o sistema tem a solução seguinte, em forma de série,

$$Y(x) = \left(I + \sum_{k=1}^{\infty} \frac{(-1)^k x^{2k} A^k}{(2k)!} \right) B + \left(xI + \sum_{k=1}^{\infty} \frac{(-1)^k x^{2k+1} A^k}{(2k+1)!} \right) C$$

convergente para $-\infty < x < \infty$.

7.21. Demonstração do teorema de existência pelo método das aproximações sucessivas

Vamos ocupar-nos em seguida da demonstração da existência e unicidade de uma solução para qualquer sistema linear e homogêneo.

$$(7.56) \quad Y'(t) = A(t)Y(t),$$

onde $A(t)$ é uma função matricial $n \times n$, contínua num intervalo aberto J . Provaremos que para qualquer ponto a em J e qualquer vector inicial B dado existe precisamente uma solução $Y(t)$ em J que satisfaz a condição inicial $Y(a) = B$.

Utilizaremos o *método das aproximações sucessivas*, um método iterativo aplicável igualmente em muitos outros problemas. O método foi publicado pela primeira vez por Liouville em 1838 em ligação com o estudo das equações diferenciais lineares de segunda ordem. Foi mais tarde estendido ao estudo de equações lineares de ordem n respectivamente por J. Caqué em 1864, L. Fuchs em 1870, e G. Reano em 1888. Em 1890 Emile Picard (1856-1941) generalizou o método de modo a abarcar também as equações diferenciais não lineares. Em reconhecimento às suas contribuições neste domínio, alguns autores citam-no como o *método de Picard*. O interesse do método é não somente teórico, mas também se usa em alguns casos para obter as aproximações numéricas das soluções.

O método começa com uma suposta solução da equação (7.56). Toma-se como tal o vector inicial B dado, embora tal não seja indispensável. Substituímos em seguida esta solução no segundo membro da equação e obtemos uma nova equação diferencial,

$$Y'(t) = A(t)B.$$

Nesta equação o segundo membro não contém já a função incógnita, pelo que a equação se pode resolver de imediato pela integração de ambos os membros de a a x , sendo x um ponto arbitrário de J . Esta equação tem precisamente uma solução Y_1 em J satisfazendo à condição inicial $Y_1(a) = B$, nomeadamente

$$Y_1(x) = B + \int_a^x A(t)B \, dt.$$

Substituamos agora $Y(t)$ por $Y_1(t)$ no segundo membro da equação diferencial original (7.56) para obtermos uma nova equação diferencial

$$Y'(t) = A(t)Y_1(t).$$

Esta equação tem uma solução única Y_2 em J com $Y_2(a) = B$,

$$(7.57) \quad Y_2(x) = B + \int_a^x A(t)Y_1(t) \, dt.$$

Substituímos então Y_2 no segundo membro de (7.56) e resolvemos a equação resultante para determinar Y_3 com $Y_3(a) = B$, e assim sucessivamente. Este processo gera uma sucessão de funções Y_0, Y_1, Y_2, \dots , onde $Y_0 = B$ e em que Y_{k+1} se determina a partir de Y_k mediante a fórmula de recorrência.

$$(7.58) \quad Y_{k+1}(x) = B + \int_a^x A(t)Y_k(t) \, dt \quad \text{para } k = 0, 1, 2, \dots$$

O nosso objectivo consiste em provar que a sucessão de funções assim definida converge para uma função limite Y , a qual é uma solução da equação diferencial (7.56) em J e satisfaz também à condição inicial $Y(a) = B$. As funções Y_0, Y_1, Y_2, \dots , chamam-se *aproximações sucessivas* de Y . Antes de estudarmos a convergência do processo ilustramo-lo com um exemplo.

EXEMPLO. Consideremos o problema de valor inicial $Y'(t) = AY(t)$, $Y(0) = B$, onde A é uma matriz constante $n \times n$. Sabemos que a solução é dada pela fórmula $Y(x) = e^{xA}B$ para todo x real. Vamos agora mostrar como pode esta solução obter-se pelo método das aproximações sucessivas.

A hipótese inicial é $Y_0(x) = B$. A fórmula de recorrência (7.58) dá-nos

$$Y_1(x) = B + \int_0^x AB \, dt = B + xAB,$$

$$Y_2(x) = B + \int_0^x AY_1(t) \, dt = B + \int_0^x (AB + tA^2B) \, dt = B + xAB + \frac{1}{2}x^2A^2B.$$

Por indução encontramos

$$Y_k(x) = B + xAB + \frac{1}{2}x^2A^2B + \dots + \frac{1}{k!}x^kA^kB = \left(\sum_{r=0}^k \frac{(xA)^r}{r!} \right) B.$$

A soma do segundo membro é parte da série de e^{xA} . Portanto, quando $k \rightarrow \infty$, obtemos

$$\lim_{k \rightarrow \infty} Y_k(x) = e^{xA}B$$

para todo x . Assim, neste exemplo pudemos mostrar directamente que as sucessivas aproximações convergem para uma solução do problema de valor inicial em $(-\infty, +\infty)$.

Demonstração da convergência da sucessão das sucessivas aproximações. Voltamos agora à sucessão geral definida pela fórmula de recorrência (7.58). Para provar que a sucessão converge escrevemos cada termo $y_k(x)$ na forma

$$(7.59) \quad Y_k(x) = Y_0(x) + \sum_{m=0}^{k-1} \{Y_{m+1}(x) - Y_m(x)\}.$$

Para provar que $Y_k(x)$ tende para um limite quando $k \rightarrow \infty$ demonstramos que a série

$$(7.60) \quad \sum_{m=0}^{\infty} \{Y_{m+1}(x) - Y_m(x)\}$$

converge para todo x em J . Com esta finalidade basta provar que a série

$$(7.61) \quad \sum_{m=0}^{\infty} \|Y_{m+1}(x) - Y_m(x)\|$$

converge. Nesta série recorreremos à norma matricial introduzida na Secção 7.3; a norma de uma matriz é a soma dos valores absolutos de todos os seus elementos.

Consideremos um subintervalo limitado e fechado J_1 de J e que contenha a . Vamos demonstrar que, para cada x em J_1 , a série em 7.61 é majorada por uma série convergente de termos constantes. Isto implica que a série converge *uniformemente* em J_1 .

Para estimar a grandeza dos termos de (7.61) aplicamos a fórmula de recorrência repetidamente. Inicialmente, temos

$$Y_1(x) - Y_0(x) = \int_a^x A(t)B \, dt.$$

Para maior simplicidade, admitimos que $a < x$. Então podemos escrever

$$(7.62) \quad \|Y_1(x) - Y_0(x)\| = \left\| \int_a^x A(t)B \, dt \right\| \leq \int_a^x \|A(t)\| \|B\| \, dt.$$

Uma vez que cada elemento de $A(t)$ é contínuo em J , cada elemento é limitado no intervalo fechado e limitado J_1 . Consequentemente $\|A(t)\| \leq M$, onde M é a soma dos limites de todos os elementos de $A(t)$ no intervalo J_1 . (O número M depende de J_1). Consequentemente o integrando em (7.62) é limitado por $\|B\| M$, pelo que temos

$$\|Y_1(x) - Y_0(x)\| \leq \int_a^x \|B\| M dt = \|B\| M(x - a)$$

para todo $x > a$ em J_1 .

Usemos agora a fórmula de recorrência, uma vez mais, para exprimir a diferença $Y_2 - Y_1$ por intermédio de $Y_1 - Y_0$, e usemos depois a estimativa acabada de obter para $Y_1 - Y_0$ para obtermos

$$\begin{aligned} \|Y_2(x) - Y_1(x)\| &= \left\| \int_a^x A(t) \{Y_1(t) - Y_0(t)\} dt \right\| \leq \int_a^x \|A(t)\| \|B\| M(t - a) dt \\ &\leq \|B\| M^2 \int_a^x (t - a) dt = \|B\| \frac{M^2(x - a)^2}{2!} \end{aligned}$$

para todo $x > a$ em J_1 . Por indução encontramos

$$\|Y_{m+1}(x) - Y_m(x)\| \leq \|B\| \frac{M^{m+1}(x - a)^{m+1}}{(m + 1)!} \quad \text{para } m = 0, 1, 2, \dots,$$

e para todo $x > a$ em J_1 . Se $x < a$ um raciocínio semelhante dá-nos a mesma desigualdade com $|x - a|$ em vez de $(x - a)$. Se representamos por L a amplitude do intervalo J_1 , então temos $|x - a| \leq L$ para todo x em J_1 , onde resulta a estimativa

$$\|Y_{m+1}(x) - Y_m(x)\| \leq \|B\| \frac{M^{m+1}L^{m+1}}{(m + 1)!} \quad \text{para } m = 0, 1, 2, \dots,$$

e para todo x em J_1 . Por conseguinte a série (7.61) é majorada pela série convergente

$$\|B\| \sum_{m=0}^{\infty} \frac{(ML)^{m+1}}{(m + 1)!} = \|B\| (e^{ML} - 1).$$

Isto demonstra que a série (7.61) converge uniformemente em J_1 .

O raciocínio efectuado mostra que a sucessão das sucessivas aproximações converge sempre e a convergência é uniforme em J_1 . Seja Y a função limite, isto é, define-se $Y(x)$ para cada x em J_1 pela equação

$$Y(x) = \lim_{k \rightarrow \infty} Y_k(x).$$

Vamos provar que Y goza das seguintes propriedades:

(a) Y é contínua em J_1 .

(b) $Y(x) = B + \int_a^x A(t) Y(t) dt$ para todo x em J_1 .

(c) $Y(a) = B$ e $Y'(x) = A(x)Y(x)$ para todo x em J_1 .

A alínea (c) mostra que Y é uma solução do problema de valor inicial em J_1 .

Demonstração de (a). Cada função Y_k é uma matriz coluna cujos elementos são funções escalares contínuas em J_1 . Cada elemento da função limite Y é o limite de uma sucessão uniformemente convergente de funções contínuas pelo que, pelo teorema 11.1 do Volume I, cada elemento de Y é também contínuo em J_1 . Consequentemente Y é ela própria contínua em J_1 .

Demonstração de (b). A fórmula de recorrência (7.58) estabelece que

$$Y_{k+1}(x) = B + \int_a^x A(t)Y_k(t) dt.$$

Deste modo

$$\begin{aligned} Y(x) &= \lim_{k \rightarrow \infty} Y_{k+1}(x) = B + \lim_{k \rightarrow \infty} \int_a^x A(t)Y_k(t) dt = B + \int_a^x A(t) \lim_{k \rightarrow \infty} Y_k(t) dt \\ &= B + \int_a^x A(t)Y(t) dt. \end{aligned}$$

A permutação do símbolo limite com o símbolo de integral é válida devido à convergência uniforme da sucessão $\{Y_k\}$ em J_1 .

Demonstração de (c). A equação $Y(a) = B$ resulta de (b). Devido a (a) o integrando em (b) é contínuo em J_1 pelo que, pelo primeiro teorema fundamental do cálculo, $Y'(x)$ existe e é igual a $A(x)Y(x)$ em J_1 .

O intervalo J_1 sabe-se que é qualquer subintervalo fechado e limitado de J contendo a . Se J_1 se aumenta, o processo de obtenção de $Y(x)$ não varia porque só inclui a integração de a e x . Posto que para todo x de J existe um subintervalo de J fechado e limitado e que contém a e x , existe uma solução em todo o intervalo J .

TEOREMA 7.17. TEOREMA DE UNICIDADE PARA SISTEMAS LINEARES HOMOGÊNEOS. Se $A(t)$ é contínuo num intervalo aberto J , a equação diferencial

$$Y'(t) = A(t)Y(t)$$

tem quando muito uma solução em J satisfazendo a uma condição inicial dada $Y(a) = B$.

Demonstração. Sejam Y e Z duas soluções em J . Seja J_1 qualquer subintervalo limitado e fechado de J , contendo a . Vamos demonstrar que $Z(x) = Y(x)$ para todo x em J_1 , o que implica que $Z = Y$ em todo intervalo J .

Visto que quer Y quer Z são soluções, temos

$$Z'(t) - Y'(t) = A(t)(Z(t) - Y(t)).$$

Escolhamos x em J_1 e integremos esta equação de a a x para obtermos

$$Z(x) - Y(x) = \int_a^x A(t)\{Z(t) - Y(t)\} dt.$$

Isto implica a desigualdade

$$(7.63) \quad \|Z(x) - Y(x)\| \leq M \left| \int_a^x \|Z(t) - Y(t)\| dt \right|,$$

onde M é um limite superior para $\|A(t)\|$ em J_1 . Seja M_1 um limite superior para função contínua $\|Z(t) - Y(t)\|$ em J_1 . Então a desigualdade (7.63) dá-nos

$$(7.64) \quad \|Z(x) - Y(x)\| \leq MM_1 |x - a|.$$

Introduzindo (7.64) no segundo membro de (7.63) obtemos

$$\|Z(x) - Y(x)\| \leq M^2 M_1 \left| \int_a^x |t - a| dt \right| = M^2 M_1 \frac{|x - a|^2}{2}.$$

Por indução encontramos

$$(7.65) \quad \|Z(x) - Y(x)\| \leq M^m M_1 \frac{|x - a|^m}{m!}.$$

Quando $m \rightarrow \infty$ o segundo membro tende para 0, pelo que $Z(x) = Y(x)$, e o teorema está demonstrado.

O resultado desta secção pode resumir-se no seguinte teorema de existência e unicidade.

TEOREMA 7.18. *Seja A uma função matricial $n \times n$, contínua num intervalo aberto J . Se $a \in J$ e B é um qualquer vector n -dimensional, o sistema linear e homogéneo.*

$$Y'(t) = A(t)Y(t), \quad Y(a) = B,$$

tem uma e uma só solução em J , um vector n -dimensional.

7.22. O método das aproximações sucessivas aplicado a sistemas não lineares de primeira ordem

O método das aproximações sucessivas pode também aplicar-se a certos sistemas não lineares. Consideremos um sistema de primeira ordem da forma

$$(7.66) \quad Y' = F(t, Y),$$

onde F é uma dada função vectorial n -dimensional, e Y é uma função vectorial n -dimensional a determinar. Pretendemos uma solução Y que verifique a equação

$$Y'(t) = F[t, Y(t)]$$

para cada t em algum intervalo J e que também satisfaz a uma condição inicial dada, seja $Y(a) = B$, onde $a \in J$ e B é um vector n -dimensional dado.

De maneira análoga ao caso linear, construímos uma sucessão das sucessivas aproximações Y_0, Y_1, Y_2, \dots , fazendo $Y_0 = B$ e definindo Y_{k+1} em função de Y_k pela fórmula de recorrência

$$(7.67) \quad Y_{k+1}(x) = B + \int_a^x F[t, Y_k(t)] dt \quad \text{para } k = 0, 1, 2, \dots$$

Sob certas condições impostas a F , esta sucessão convergirá para uma função limite Y que satisfará à equação diferencial dada e à condição inicial dada.

Antes de analisarmos a convergência do processo vamos analisar a resolução de alguns exemplos unidimensionais, escolhidos para ilustrar algumas das dificuldades que podem apresentar-se na prática.

EXEMPLO 1. Consideremos o problema de valor inicial não linear $y' = x^2 + y^2$, com $y = 0$ quando $x = 0$. Calculemos algumas aproximações da solução. Escolhemos $Y_0(x) = 0$ e determinamos as três aproximações seguintes como se indica:

$$Y_1(x) = \int_0^x t^2 dt = \frac{x^3}{3},$$

$$Y_2(x) = \int_0^x [t^2 + Y_1^2(t)] dt = \int_0^x \left(t^2 + \frac{t^6}{9} \right) dt = \frac{x^3}{3} + \frac{x^7}{63},$$

$$Y_3(x) = \int_0^x \left[t^2 + \left(\frac{t^3}{3} + \frac{t^7}{63} \right)^2 \right] dt = \frac{x^3}{3} + \frac{x^7}{63} + \frac{2x^{11}}{2079} + \frac{x^{15}}{59535}.$$

É agora claro que será necessário um enorme trabalho para calcular outras aproximações. Por exemplo, as duas aproximações seguintes Y_4 e Y_5 serão polinómios de graus 31 e 63, respectivamente.

O exemplo que apresentamos a seguir apresenta uma dificuldade suplementar que pode surgir no cálculo das sucessivas aproximações.

EXEMPLO 2. Consideremos o problema de valor inicial não linear $y' = 2x + e^y$, com $y = 0$ quando $x = 0$. Partimos da hipótese inicial $Y_0(x) = 0$ e calculamos

$$Y_1(x) = \int_0^x (2t + 1) dt = x^2 + x,$$

$$Y_2(x) = \int_0^x (2t + e^{t^2+t}) dt = x^2 + \int_0^x e^{t^2+t} dt.$$

Aqui está-se impedido de presseguir pelo facto de que o último integral não pode ser calculado por meio de funções elementares. Porém, para um dado x é possível calcu-

lar uma aproximação numérica para o integral e, por seu intermédio, obter uma aproximação para $Y_2(x)$.

Como o provam as dificuldades mostradas nos dois exemplos, o método das aproximações sucessivas não é, por vezes, muito útil para a determinação explícita de soluções. O valor real do método reside na sua aplicação ao estabelecimento de teoremas de existência.

7.23. Demonstração de um teorema de existência e unicidade para sistemas não lineares de primeira ordem

Voltamos agora a nossa atenção para um teorema de existência e unicidade aplicável a sistemas de primeira ordem não lineares. Impondo restrições adequadas à função que aparece no segundo membro da equação diferencial

$$Y' = F(x, Y),$$

podemos generalizar o método de demonstração usado no caso linear da Secção 7.21.

Seja J o intervalo aberto onde se pretende uma solução. Admitamos $a \in J$ e seja B um vector n -dimensional dado. Representamos por S um conjunto no espaço de dimensão $(n + 1)$ definido por

$$S = \{(x, Y) \mid |x - a| \leq h, \|Y - B\| \leq k\},$$

onde $h > 0$ e $k > 0$. [Se $n = 1$, trata-se de um rectângulo com centro em (a, B) e cujos lados medem $2h$ e $2k$]. Admitamos que o domínio de F inclui um conjunto S deste tipo e que F é limitado em S , isto é,

$$(7.68) \quad \|F(x, Y)\| \leq M$$

para todo (x, Y) em S , sendo M uma constante positiva.

Seguidamente admitamos que a função composta $G(x) = F(x, Y(x))$ é contínua no intervalo $(a - h, a + h)$ para toda a função Y que seja contínua em $(a - h, a + h)$ e que goza da propriedade de $(x, Y(x)) \in S$ para todo o x em $(a - h, a + h)$. Esta hipótese assegura a existência dos integrais que apareçam no método de aproximações sucessivas e implica também a continuidade das funções assim construídas.

Finalmente, admitamos que F verifica uma condição da forma

$$\|F(x, Y) - F(x, Z)\| \leq A \|Y - Z\|$$

— para todo o par de pontos (x, Y) e (x, Z) em S , sendo A uma constante positiva. Esta é a chamada *condição de Lipschitz*, em homenagem a Rudolph Lipschitz que foi o primeiro a introduzi-la em 1876. Uma condição de Lipschitz não é uma restrição muito forte para uma dada função e permite-nos, em contrapartida, generalizar a demonstração da existência e unicidade do caso linear para o caso não linear.

TEOREMA 7.19. EXISTÊNCIA E UNICIDADE DE SOLUÇÕES PARA SISTEMAS DE PRIMEIRA ORDEM NÃO LINEARES. *Admita-se que F é limitada e contínua e que satisfaz, no conjunto S , às condições de Lipschitz atrás referidas. Se I é um intervalo aberto $(a - c, a + c)$, onde $c = \min \{h, k/M\}$ então existe uma e uma só função Y definida em I , com $Y(a) = B$, tal que $(x, Y(x)) \in S$ e*

$$Y'(x) = F(x, Y(x)) \quad \text{para todo } x \text{ em } I.$$

Demonstração. Visto a demonstração ser análoga à do caso linear, indicamos unicamente os passos principais. Seja $Y_0(x) = B$ e definamos as funções vectoriais Y_1, Y_2, \dots , em I pela fórmula de recorrência.

$$(7.69) \quad Y_{m+1}(x) = B + \int_a^x F[t, Y_m(t)] dt \quad \text{para } m = 0, 1, 2, \dots$$

Para que a fórmula de recorrência seja provida de significado necessitamos saber que $(x, Y_m(x)) \in S$ para todo x em I . Isto é facilmente demonstrado por indução em m . Quando $m = 0$ temos $(x, Y_0(x)) = (x, B)$, que está em S . Admitamos então que $(x, Y_m(x)) \in S$ para algum m e cada x de I . Utilizando (7.69) e (7.68) obtemos

$$\|Y_{m+1}(x) - B\| \leq \left| \int_a^x \|F[t, Y_m(t)]\| dt \right| \leq M \left| \int_a^x dt \right| = M|x - a|.$$

Visto ser $|x - a| \leq c$ para todo x em I , isto implica que

$$\|Y_{m+1}(x) - B\| \leq Mc \leq k,$$

o que mostra que $(x, Y_{m+1}(x)) \in S$ para todo x em I . Consequentemente a fórmula de recorrência é provida de significado para todo $m \geq 0$ e todo x em I .

A convergência da sucessão $\{Y_m(x)\}$ pode agora estabelecer-se exactamente como na Secção 7.21. Escrevamos

$$Y_k(x) = Y_0(x) + \sum_{m=0}^{k-1} \{Y_{m+1}(x) - Y_m(x)\}$$

e provemos que $Y_k(x)$ tende para um limite quando $k \rightarrow \infty$, demonstrando que a série

$$\sum_{m=0}^{\infty} \|Y_{m+1}(x) - Y_m(x)\|$$

converge em I . Isto pode concluir-se da desigualdade

$$\|Y_{m+1}(x) - Y_m(x)\| \leq \frac{MA^m |x - a|^{m+1}}{(m+1)!} \leq \frac{MA^m c^{m+1}}{(m+1)!}$$

que se demonstra por indução, usando a fórmula de recorrência e a condição de Lipschitz. Definimos então a função limite Y por

$$Y(x) = \lim_{m \rightarrow \infty} Y_m(x)$$

para todo x em I e verificamos que satisfaz a equação integral

$$Y(x) = B + \int_a^x F[t, Y(t)] dt,$$

precisamente como no caso linear. Está assim demonstrada a existência da solução. A unicidade pode então demonstrar-se pelo mesmo método utilizado na demonstração do teorema 7.17.

7.24. Exercícios

1. Considerar o problema de valor inicial linear

$$y' + y = 2e^x, \quad \text{com } y = 1 \text{ quando } x = 0.$$

(a) Determinar a solução exacta Y deste problema.

(b) Aplicar o método das aproximações sucessivas, partindo da hipótese inicial $Y_0(x) = 1$. Determinar $Y_n(x)$ explicitamente e mostrar que

$$\lim_{n \rightarrow \infty} Y_n(x) = Y(x)$$

para todo real x .

2. Aplicar o método das aproximações sucessivas ao problema de valor inicial não linear

$$y' = x + y^2, \quad \text{com } y = 0 \text{ quando } x = 0.$$

Tomar $Y_0(x) = 0$ como função inicial e calcular $Y_3(x)$.

3. Aplicar o método de aproximações sucessivas ao problema de valor inicial não linear

$$y' = 1 + xy^2, \quad \text{com } y = 0 \text{ quando } x = 0.$$

Tomar $Y_0(x) = 0$ como função inicial e calcular $Y_3(x)$.

4. Aplicar o método das aproximações sucessivas ao problema de valor inicial não linear

$$y' = x^2 + y^2, \quad \text{com } y = 0 \text{ quando } x = 0.$$

Considerar a "inconveniente" hipótese inicial $Y_0(x) = 1$, calcular $Y_3(x)$, e comparar com o resultado do Exemplo 1 da Secção 7.22.

5. Considerar o problema de valor inicial não linear

$$y' = x^2 + y^2, \quad \text{com } y = 1 \text{ quando } x = 0.$$

(a) Aplicar o método das aproximações sucessivas, partindo da hipótese inicial $Y_0(x) = 1$ e calcular $Y_2(x)$.

(b) Seja $R = [-1, 1] \times [-1, 1]$. Determinar o menor M tal que $|f(x, y)| \leq M$ em R . Achar um intervalo $I = (-c, c)$ tal que o gráfico de cada função de aproximação Y_n em I esteja em R .

(c) Supor que a solução $y = Y(x)$ admite um desenvolvimento em série de potências na vizinhança da origem. Determinar os seis primeiros termos não nulos deste desenvolvimento e comparar com o resultado da alínea (a).

6. Considerar o problema de valor inicial

$$y' = 1 + y^2, \quad \text{com } y = 0 \text{ quando } x = 0.$$

(a) Aplicar o método das aproximações sucessivas, partindo da hipótese inicial $Y_0(x) = 0$ e calcular $Y_4(x)$.

(b) Provar que toda a função de aproximação Y_n está definida em todo o eixo real.

(c) Usar o teorema 7.19 para mostrar que o problema de valor inicial tem, quando muito, uma solução em qualquer intervalo da forma $(-h, h)$.

(d) Resolver a equação diferencial por separação de variáveis e demonstrar assim que existe uma única solução Y do problema de valor inicial no intervalo $(-\pi/2, \pi/2)$ e nunca num intervalo mais amplo. Neste exemplo, as aproximações sucessivas definem-se em todo o eixo real, mas convergem para uma função limite unicamente no intervalo $(-\pi/2, \pi/2)$.

7. Determinar duas funções $y = Y(x)$ e $z = Z(x)$ que satisfaçam simultaneamente ao sistema de equações.

$$y' = z, \quad z' = x^3(y + z)$$

com as condições iniciais $y = 1$ e $z = 1/2$ quando $x = 0$. Partir da hipótese inicial $Y_0(x) = 1$, $Z_0(x) = 1/2$ e utilizar o método das aproximações sucessivas para obter as funções de aproximação

$$Y_3(x) = 1 + \frac{x}{2} + \frac{3x^5}{40} + \frac{x^6}{60} + \frac{x^9}{192},$$

$$Z_3(x) = \frac{1}{2} + \frac{3x^4}{8} + \frac{x^5}{10} + \frac{3x^8}{64} + \frac{7x^9}{360} + \frac{x^{12}}{256}.$$

8. Considerar o sistema de equações

$$y' = 2x + z, \quad z' = 3xy + x^2z,$$

com condições iniciais $y = 2$ e $z = 0$ quando $x = 0$. Partindo das hipóteses iniciais $Y_0(x) = 2$, $Z_0(x) = 0$, utilizar o método das aproximações sucessivas e determinar $Y_3(x)$ e $Z_3(x)$.

9. Considerar o problema de valor inicial

$$y'' = x^2y' + x^4y, \quad \text{com } y = 5 \quad \text{e} \quad y' = 1 \text{ quando } x = 0.$$

Transformar este problema noutro equivalente que incluía um sistema de duas equações com duas funções desconhecidas $y = Y(x)$ e $z = Z(x)$, sendo $z = y'$. Aplicar então o método das aproximações sucessivas, partindo das funções iniciais $Y_0(x) = 5$ e $Z_0(x) = 1$ e determinar $Y_3(x)$ e $Z_3(x)$.

10. Seja f uma função definida no rectângulo $R = [-1, 1] \times [-1, 1]$ do seguinte modo:

$$f(x, y) = \begin{cases} 0 & \text{if } x = 0, \\ 2y/x & \text{if } x \neq 0 \text{ e } |y| \leq x^2, \\ 2x & \text{if } x \neq 0 \text{ e } y > x^2, \\ -2x & \text{if } x \neq 0 \text{ e } y < -x^2. \end{cases}$$

- (a) Provar que $|f(x, y)| \leq 2$ para todo (x, y) em R .
- (b) Mostrar que f não satisfaz em R à condição de Lipchitz.
- (c) Para cada constante C satisfazendo a $|C| \leq 1$, mostrar que $y = Cx^2$ é uma solução do problema de valor inicial $y' = f(x, y)$, com $y = 0$ quando $x = 0$. Mostrar também que o gráfico de cada uma destas soluções em $(-1, 1)$ está em R .
- (d) Aplicar o método das aproximações a este problema de valor inicial, partindo com a função inicial $Y_0(x) = 0$. Determinar $Y_n(x)$ e mostrar que as aproximações convergem para uma solução do problema no intervalo $(-1, 1)$.
- (e) Repetir a alínea (d), partindo da função inicial $Y_0(x) = x$. Determinar $Y_n(x)$ e mostrar que as funções de aproximação convergem para uma solução diferente de qualquer das soluções da alínea (c).
- (f) Repetir a alínea (d), partindo com a função inicial $Y_0(x) = x^3$.
- (g) Repetir a alínea (d), partindo com a função inicial $Y_0(x) = x^{1/2}$.

★ 7.25. Aproximações sucessivas e pontos fixos de operadores

A ideia fundamental do método das aproximações sucessivas pode utilizar-se, não somente para estabelecer teoremas de existência para as equações diferenciais, mas também em muitas outras importantes questões de análise. O resto do presente capítulo reformula o método das aproximações sucessivas de uma maneira que amplia muito o âmbito das suas aplicações.

Na demonstração do teorema 7.18 construímos uma sucessão de funções $\{Y_k\}$ segundo a fórmula

$$Y_{k+1}(x) = B + \int_a^x AY_k(t) dt.$$

O segundo membro desta fórmula pode considerar-se como um operador T que converte certas funções Y em novas funções $T(Y)$ mediante a equação

$$T(Y) = B + \int_a^x AY(t) dt.$$

Na demonstração do teorema 7.8 encontramos que a solução Y do problema de valor inicial $Y'(t) = AY(t)$, $Y(a) = B$, satisfaz à equação integral

$$Y = B + \int_a^x AY(t) dt.$$

Na notação por intermédio do operador T , a igualdade anterior significa que $Y = T(Y)$. Por outras palavras, a solução Y não se altera por aplicação do operador T . Uma tal função Y diz-se um *ponto fixo* do operador T .

Vários problemas importantes de Análise se podem formular de modo que a sua solução dependa da existência de um ponto fixo para certo operador. Por conseguinte vale a pena tentar descobrir propriedades de operadores que garantam a existência de um ponto fixo. Vamos por tal motivo passar ao estudo sistemático de tal problema.

★ 7.26. Espaços lineares normados

Para formular o método das aproximações sucessivas de uma maneira geral é conveniente trabalhar com espaços lineares. Seja S um espaço linear arbitrário. Quando falamos de aproximação de um elemento x de S mediante outro elemento y de S , consideramos a diferença $x - y$, a qual se designa por *erro* de aproximação. Para medir a dimensão deste erro introduzimos uma norma no espaço.

DEFINIÇÃO DE NORMA. *Seja S um espaço linear qualquer. Uma função real N definida em S chama-se norma se possuir as seguintes propriedades:*

- (a) $N(x) \geq 0$ para todo x de S .
- (b) $N(cx) = |c| N(x)$ para todo x de S e qualquer escalar c .
- (c) $N(x + y) \leq N(x) + N(y)$ para todo x e y de S .
- (d) $N(x) = 0$ implica $x = O$.

Um espaço linear ao qual se atribui determinada norma diz-se um *espaço linear normado*.

A norma de x representa-se por vezes por $\|x\|$ em vez de $N(x)$. Nesta notação as propriedades fundamentais escrevem-se:

- (a) $\|x\| \geq 0$ para todo x em S .
- (b) $\|cx\| = |c| \|x\|$ para todo x em S e qualquer escalar c .
- (c) $\|x + y\| \leq \|x\| + \|y\|$ para todo (x, y) de S .
- (d) $\|x\| = 0$ implica $x = O$.

Se x e y pertencem a S , designa-se $\|x - y\|$ como sendo a *distância* de x a y .

Se o espaço é euclidiano, então possui sempre uma norma que é consequência do produto interno, isto é, $\|x\| = (x, x)^{1/2}$. Contudo vamos interessar-nos por uma norma particular, a qual não resulta da definição de produto interno.

EXEMPLO. *A norma max.* Seja $C(J)$ o espaço linear das funções reais contínuas num intervalo fechado e limitado J . Se $\varphi \in C(J)$, definimos

$$\|\varphi\| = \max_{x \in J} |\varphi(x)|,$$

onde o símbolo do segundo membro significa o máximo valor absoluto de φ em J . O leitor pode verificar que esta norma goza das quatro propriedades fundamentais.

A norma max não resulta da definição de produto interno. Para demonstrar tal afirmação basta provar que a norma max não possui alguma das propriedades a que satisfazem todas as normas deduzidas a partir da noção de produto interno. Por exemplo, se uma norma resulta de um produto interno, “a regra do paralelogramo”

$$\|x + y\|^2 + \|x - y\|^2 = 2 \|x\|^2 + 2 \|y\|^2$$

é válida para todo o par x e y de S . (Ver Exercício 16 da Secção 1.13.) A regra do paralelogramo nem sempre é satisfeita pela norma max. Por exemplo, sejam x e y funções definidas no intervalo $[0, 1]$ por

$$x(t) = t, \quad y(t) = 1 - t.$$

Então temos $\|x\| = \|y\| = \|x + y\| = \|x - y\| = 1$, pelo que a regra do paralelogramo não se verifica.

★ 7.27. Operadores de contracção

Nesta Secção consideramos o espaço linear normado $C(J)$ de todas as funções reais contínuas num intervalo fechado e limitado J , no qual $\|\varphi\|$ é a norma max. Consideremos um operador

$$T: C(J) \rightarrow C(J)$$

de domínio $C(J)$ e cujo contradomínio é um subconjunto de $C(J)$, quer dizer se φ é contínua em J , então $T(\varphi)$ também é contínua em J . As fórmulas seguintes representam alguns exemplos simples de tais operações. Em cada caso φ é uma função arbitrária em $C(J)$ e $T(\varphi)(x)$ define-se para cada x de J pela fórmula que se indica

$$T(\varphi)(x) = \lambda \varphi(x), \quad \text{onde } \lambda \text{ é um número real fixo,}$$

$$T(\varphi)(x) = \int_c^x \varphi(t) dt, \quad \text{onde } c \text{ é um dado ponto de } J.$$

$$T(\varphi)(x) = b + \int_c^x f[t, \varphi(t)] dt,$$

onde b é uma constante e a função composta $f[t, \varphi(t)]$ é contínua em J .

Interessam-nos agora aqueles operadores T para os quais a distância $\|T(\varphi) - T(\psi)\|$ é menos do que o produto de $\|\varphi - \psi\|$ por uma constante $\alpha < 1$. Estes são os chamados *operadores de contracção*; definem-se de modo seguinte:

DEFINIÇÃO DE UM OPERADOR DE CONTRACÇÃO. Um operador $T: C(J) \rightarrow C(J)$ diz-se de contracção se existir uma constante α , verificando $0 \leq \alpha < 1$, tal que para todo o par de funções φ e ψ de $C(J)$ se tenha

$$(7.70) \quad \|T(\varphi) - T(\psi)\| \leq \alpha \|\varphi - \psi\|.$$

A constante α chama-se constante de contracção para T .

Nota: A desigualdade (7.70) é válida se e só se

$$|T(\varphi)(x) - T(\psi)(x)| \leq \alpha \|\varphi - \psi\| \quad \text{para qualquer } x \text{ de } J.$$

EXEMPLO 1. Seja T um operador definido por $T(\varphi)(x) = \lambda\varphi(x)$, onde λ é uma constante. Visto que

$$|T(\varphi)(x) - T(\psi)(x)| = |\lambda| |\varphi(x) - \psi(x)|,$$

tem-se $\|T(\varphi) - T(\psi)\| = |\lambda| \|\varphi - \psi\|$. Consequentemente este é um operador de contracção se e só se $|\lambda| < 1$, caso em que $|\lambda|$ pode usar-se como constante de contracção.

EXEMPLO 2. Seja $T(\varphi)(x) = b + \int_c^x f[t, \varphi(t)] dt$, onde f satisfaz à condição de Lipschitz da forma

$$|f(x, y) - f(x, z)| \leq K|y - z|$$

para todo x de J e quaisquer reais z e y ; K é uma constante positiva. Seja $L(J)$ a amplitude do intervalo J . Se $KL(J) < 1$, podemos facilmente provar que T é um operador de contracção $KL(J)$. Com efeito, para todo x em J temos

$$\begin{aligned} |T(\varphi)(x) - T(\psi)(x)| &= \left| \int_c^x \{f[t, \varphi(t)] - f[t, \psi(t)]\} dt \right| \leq K \left| \int_c^x |\varphi(t) - \psi(t)| dt \right| \\ &\leq K \|\varphi - \psi\| \left| \int_c^x dt \right| \leq KL(J) \|\varphi - \psi\|. \end{aligned}$$

Se $KL(J) < 1$, então T é um operador de contracção com constante de contracção $\alpha = KL(J)$.

★ 7.28. Teorema do ponto fixo para operadores de contracção

O teorema seguinte mostra que cada operador de contracção admite um único ponto fixo.

TEOREMA 7.20. Se $T: C(J) \rightarrow C(J)$ é um operador de contracção então existe uma e uma só função φ em $C(J)$ tal que

$$(7.71) \quad T(\varphi) = \varphi.$$

Demonstração. Seja φ_0 qualquer função de $C(J)$ e definamos uma sucessão de funções $\{\varphi_n\}$ mediante a fórmula de recorrência

$$\varphi_{n+1} = T(\varphi_n) \quad \text{para } n = 0, 1, 2, \dots$$

Observe-se que $\varphi_{n+1} \in C(J)$ para cada n . Demonstraremos que a sucessão $\{\varphi_n\}$ converge para uma função limite φ em $C(J)$. O método é análogo ao utilizado na demonstração do teorema 7.18. Escrevemos cada φ_n como uma soma da forma

$$(7.72) \quad \varphi_n(x) = \varphi_0(x) + \sum_{k=0}^{n-1} \{\varphi_{k+1}(x) - \varphi_k(x)\}$$

e provamos a convergência de $\{\varphi_n\}$ demonstrando que a série

$$(7.73) \quad \varphi_0(x) + \sum_{k=0}^{\infty} \{\varphi_{k+1}(x) - \varphi_k(x)\}$$

converge uniformemente em J . Em seguida provaremos que a soma desta série é o ponto fixo referido.

A convergência uniforme da série estabelecer-se-á por comparação com a série geométrica convergente

$$M \sum_{k=0}^{\infty} \alpha^k,$$

onde $M = \|\varphi_0\| + \|\varphi_1\|$, e α é uma constante de contracção para T . A comparação é proporcionada pela desigualdade

$$(7.74) \quad |\varphi_{k+1}(x) - \varphi_k(x)| \leq M\alpha^k$$

a qual permanece válida para cada x em J e cada $k \geq 1$. Para demonstrar (7.74) observamos que

$$|\varphi_{k+1}(x) - \varphi_k(x)| = |T(\varphi_k)(x) - T(\varphi_{k-1})(x)| \leq \alpha \|\varphi_k - \varphi_{k-1}\|.$$

Por conseguinte a desigualdade (7.74) ficará demonstrada se provarmos que

$$(7.75) \quad \|\varphi_k - \varphi_{k-1}\| \leq M\alpha^{k-1}$$

para todo $k \geq 1$. Demonstramos (7.75) por indução. Para $k = 1$, temos

$$\|\varphi_1 - \varphi_0\| \leq \|\varphi_1\| + \|\varphi_0\| = M,$$

a qual é a mesma que (7.75). Para demonstrar que (7.75) é válida para $k + 1$ se o for para k , observemos que

$$|\varphi_{k+1}(x) - \varphi_k(x)| = |T(\varphi_k)(x) - T(\varphi_{k-1})(x)| \leq \alpha \|\varphi_k - \varphi_{k-1}\| \leq M\alpha^k.$$

Uma vez que a desigualdade anterior é válida para todo o x em J , ter-se-á

$$\|\varphi_{k+1} - \varphi_k\| \leq M\alpha^k.$$

Isto demonstra (7.75) por indução. Por conseguinte a série (7.73) converge uniformemente em J . Se representarmos a sua soma por $\varphi(x)$ temos

$$(7.76) \quad \varphi(x) = \lim_{n \rightarrow \infty} \varphi_n(x) = \varphi_0(x) + \sum_{k=0}^{\infty} \{\varphi_{k+1}(x) - \varphi_k(x)\}.$$

A função φ é contínua em J porque é a soma de uma série uniformemente convergente de funções contínuas. Para demonstrar que φ é um ponto fixo de T comparamos $T(\varphi)$ com $\varphi_{n+1} = T(\varphi_n)$. Recorrendo à propriedade de contracção de T temos

$$(7.77) \quad |T(\varphi)(x) - \varphi_{n+1}(x)| = |T(\varphi)(x) - T(\varphi_n)(x)| \leq \alpha |\varphi(x) - \varphi_n(x)|.$$

Mas de (7.72) e (7.76) encontramos

$$|\varphi(x) - \varphi_n(x)| = \left| \sum_{k=n}^{\infty} \{\varphi_{k+1}(x) - \varphi_k(x)\} \right| \leq \sum_{k=n}^{\infty} |\varphi_{k+1}(x) - \varphi_k(x)| \leq M \sum_{k=n}^{\infty} \alpha^k,$$

tendo-se utilizado (7.74) no último passo da demonstração. Por conseguinte (7.77) implica

$$|T(\varphi)(x) - \varphi_{n+1}(x)| \leq M \sum_{k=n}^{\infty} \alpha^{k+1}.$$

Quando $n \rightarrow \infty$, a série do segundo membro tende para 0, pelo que $\varphi_{n+1}(x) \rightarrow T(\varphi)(x)$. Mas visto que $\varphi_{n+1}(x) \rightarrow \varphi(x)$ quando $n \rightarrow \infty$, isto prova que $\varphi(x) = T(\varphi)(x)$ para todo x em J . Consequentemente $\varphi = T(\varphi)$, pelo que φ é um ponto fixo.

Por fim vamos demonstrar que o ponto fixo φ é único. Seja ψ outra função de $C(J)$ tal que $T(\psi) = \psi$. Então tem-se

$$\|\varphi - \psi\| = \|T(\varphi) - T(\psi)\| \leq \alpha \|\varphi - \psi\|.$$

Daqui resulta que $(1-\alpha)\|\varphi - \psi\| \leq 0$. Atendendo a que $\alpha < 1$ podemos dividir por $1-\alpha$ para obtermos a desigualdade $\|\varphi - \psi\| \leq 0$. Mas porque também se tem $\|\varphi - \psi\| \geq 0$, isto significa que $\|\varphi - \psi\| = 0$ e portanto $\varphi - \psi = 0$. Está pois completamente demonstrado o teorema do ponto fixo.

★ 7.29. Aplicações do teorema do ponto fixo.

Para se fazer uma ideia do âmbito de aplicação do teorema do ponto fixo vamos utilizá-lo para demonstrar dois teoremas importantes. O primeiro dá-nos uma condição suficiente para que uma equação da forma $f(x, y) = 0$ defina y como função de x .

TEOREMA 7.21. UM TEOREMA SOBRE A EXISTÊNCIA DA FUNÇÃO IMPLÍCITA. *Se f se define numa banda R da forma*

$$R = \{(x, y) \mid a \leq x \leq b, -\infty < y < +\infty\}.$$

se a derivada parcial $D_2 f(x, y)^{(+0)}$ existe e satisfaz à desigualdade da forma

$$(7.78) \quad 0 < m \leq D_2 f(x, y) \leq M$$

para todo (x, y) em R , sendo m e M constantes em que $m \leq M$, e se ainda para cada função φ contínua em $[a, b]$ a função composta $g(x) = f[x, \varphi(x)]$ é contínua em $[a, b]$, então existe uma e uma só função $y = Y(x)$, contínua em $[a, b]$, e tal que

$$(7.79) \quad f[x, Y(x)] = 0$$

para todo x em $[a, b]$.

Nota: Expressamos este resultado dizendo que a equação $f(x, y) = 0$ serve para definir y implicitamente como função de x em $[a, b]$.

Demonstração. Designemos por C o espaço linear das funções contínuas em $[a, b]$, e definamos um operador $T: C \rightarrow C$ pela equação

$$T(\varphi)(x) = \varphi(x) - \frac{1}{M} f[x, \varphi(x)]$$

para cada x em $[a, b]$. Aqui M é a constante positiva de (7.78). A função $T(\varphi) \in C$ sempre que $\varphi \in C$. Provaremos que T é um operador de contracção. Uma vez sabido isto deduzir-se-á que T tem um ponto fixo único Y em C . Para esta função Y temos $Y = T(Y)$, o que significa

$$Y(x) = Y(x) - \frac{1}{M} f[x, Y(x)]$$

para todo x em $[a, b]$. Isto dá-nos (7.79), como se pretendia.

Para demonstrar que T é um operador de contracção consideramos a diferença

$$(7.80) \quad T(\varphi)(x) - T(\psi)(x) = \varphi(x) - \psi(x) - \frac{f[x, \varphi(x)] - f[x, \psi(x)]}{M}.$$

De acordo com o teorema da média tem-se

$$f[x, \varphi(x)] - f[x, \psi(x)] = D_2 f[x, z(x)] [\varphi(x) - \psi(x)],$$

com $z(x)$ entre $\varphi(x)$ e $\psi(x)$. Por conseguinte (7.80) dá-nos

$$(7.81) \quad T(\varphi)(x) - T(\psi)(x) = [\varphi(x) - \psi(x)] \left(1 - \frac{D_2 f[x, z(x)]}{M} \right).$$

A hipótese (7.78) implica que

$$0 \leq 1 - \frac{D_2 f[x, z(x)]}{M} \leq 1 - \frac{m}{M}.$$

Consequentemente (7.81) dá-nos a desigualdade

$$(7.82) \quad |T(\varphi)(x) - T(\psi)(x)| \leq |\varphi(x) - \psi(x)| \left(1 - \frac{m}{M}\right) \leq \alpha \|\varphi - \psi\|,$$

onde $\alpha = 1 - m/M$. Visto que $0 < m \leq M$, temos $0 \leq \alpha < 1$. A desigualdade (7.82) é válida para todo x em $[a, b]$. Logo T é um operador de contracção, e a demonstração está completada.

A outra aplicação do teorema do ponto fixo estabelece um teorema de existência para a equação integral

$$(7.83) \quad \varphi(x) = \psi(x) + \lambda \int_a^b K(x, t)\varphi(t) dt.$$

φ é uma função dada, continua em $[a, b]$, λ é uma constante dada, e K é uma dada função definida e limitada no quadrado

$$S = \{(x, y) \mid a \leq x \leq b, a \leq y \leq b\}.$$

A função K chama-se o *núcleo* da equação integral. Seja C o espaço linear das funções contínuas em $[a, b]$. Suponhamos que o núcleo K é tal que o operador T definido por

$$T(\varphi)(x) = \psi(x) + \lambda \int_a^b K(x, t)\varphi(t) dt$$

aplica C em C . Quer dizer que $T(\varphi) \in C$ sempre que $\varphi \in C$. Uma solução da equação integral é qualquer função φ de C que satisfaz (7.83).

TEOREMA 7.22. UM TEOREMA DE EXISTÊNCIA PARA EQUAÇÕES INTEGRAIS. *Se, sob as condições precedentes, se tem*

$$(7.84) \quad |K(x, y)| \leq M$$

para todo o ponto (x, y) de S , com $M > 0$, então para cada λ tal que

$$(7.85) \quad |\lambda| < \frac{1}{M(b-a)}$$

existe uma e uma só função φ em C que satisfaz à equação integral (7.83).

Demonstração. Provemos que T é um operador de contracção. Tomemos duas quaisquer funções φ_1 e φ_2 de C e consideremos a diferença

$$T(\varphi_1)(x) - T(\varphi_2)(x) = \lambda \int_a^b K(x, t) [\varphi_1(t) - \varphi_2(t)] dt.$$

Utilizando a desigualdade (7.84) podemos escrever

$$|T(\varphi_1)(x) - T(\varphi_2)(x)| \leq |\lambda| M(b-a) \|\varphi_1 - \varphi_2\| = \alpha \|\varphi_1 - \varphi_2\|,$$

onde $\alpha = |\lambda| M(b-a)$. Por (7.85) resulta que $0 \leq \alpha < 1$, pelo que T é um operador de contracção com a constante de contracção α . Deste modo T tem um único ponto fixo φ em C . Esta função φ satisfaz a (7.83).

PARTE 2

ANÁLISE NÃO LINEAR

8

CÁLCULO DIFERENCIAL EM CAMPOS ESCALARES E VECTORIAIS

8.1. Funções de \mathbb{R}^n em \mathbb{R}^m . Campos vectoriais e escalares

A parte 1 deste volume foi dedicada principalmente às transformações lineares

$$T: V \rightarrow W$$

de um espaço linear V noutro espaço linear W . Na parte 2 vamos prescindir da exigência de que T seja linear, mas restringimos os espaços V e W a serem de dimensão finita. Concretamente, consideraremos funções com domínio no espaço n dimensional \mathbb{R}^n e contradomínio no espaço m -dimensional \mathbb{R}^m .

Quando ambos os valores m e n são iguais a 1, uma tal função diz-se uma função real de uma variável real. Quando $n = 1$ e $m > 1$ diz-se uma função vectorial de uma variável real. No Volume I foram amplamente estudados exemplos de tais funções.

No presente capítulo suporemos $n > 1$ e $m \geq 1$. Quando $m = 1$ a função diz real de uma variável vectorial ou, mais sucintamente, *um campo escalar*. Quando $m > 1$ chama-se uma função vectorial de uma variável vectorial, ou simplesmente *um campo vectorial*.

Este capítulo estende os conceitos de limite, continuidade e derivada a campos escalares e vectoriais. Nos Capítulos 10 e 11 faz-se o mesmo para o conceito de integral.

Notação: Os escalares representar-se-ão com letra de tipo corrente, e os vectores com letras tipo negrito. Se f é um campo escalar definido num ponto $\bar{x} = (x_1, \dots, x_n)$ de \mathbb{R}^n , as notações $f(\bar{x})$ e $f(x_1, x_2, \dots, x_n)$ usam-se indistintamente para representar o valor de f num certo ponto. Se f é um campo vectorial escrevemos $f(\bar{x})$ ou $f(x_1, \dots, x_n)$ para representar o valor da função em \bar{x} . Utilizaremos o produto interno

$$\mathbf{x} \cdot \mathbf{y} = \sum_{k=1}^n x_k y_k$$

e a correspondente norma $\|\mathbf{x}\| = (\mathbf{x} \cdot \mathbf{x})^{1/2}$, onde $\mathbf{x} = (x_1, x_2, \dots, x_n)$ e $\mathbf{y} = (y_1, y_2, \dots, y_n)$. Pontos de \mathbf{R}^2 representam-se habitualmente por (x, y) em vez de (x_1, x_2) ; pontos de \mathbf{R}^3 por (x, y, z) em vez de (x_1, x_2, x_3) .

Os campos escalares e vectoriais definidos em subconjuntos de \mathbf{R}^2 e \mathbf{R}^3 apresentam-se com frequência nas aplicações da matemática às Ciências e Engenharia. Por exemplo, se a cada ponto \mathbf{x} da atmosfera atribuírmos um número real $f(\mathbf{x})$ que represente a temperatura em \mathbf{x} , a função f assim definida constitui um campo escalar. Se atribuírmos a cada ponto citado num vector representando a velocidade do vento naquele ponto, obtemos um exemplo de um campo vectorial.

Em problemas de Física tratando quer com campos escalares, quer com campos vectoriais, é importante saber como varia o campo ao passar de um para outro ponto. No caso unidimensional a derivada é o instrumento matemático que dá conta dessa mudança. A teoria da derivação no caso unidimensional trata com funções definidas em intervalos abertos. Para generalizar a teoria a \mathbf{R}^n , vamos considerar generalizações de intervalos abertos chamados *conjuntos abertos*.

8.2. Bolas abertas e conjuntos abertos

Sejam \mathbf{a} um ponto dado em \mathbf{R}^n e r um número positivo dado. O conjunto de todos os pontos \mathbf{x} de \mathbf{R}^n tais que

$$\|\mathbf{x} - \mathbf{a}\| < r$$

chama-se *n-bola aberta* de raio r e centro \mathbf{a} . Representamos tal conjunto por $B(\mathbf{a})$ ou por $B(\mathbf{a}; r)$.

A bola $B(\mathbf{a}; r)$ consiste de todos os pontos cuja distância a \mathbf{a} é menor que r . Em \mathbf{R}^1 ela é muito simplesmente um intervalo aberto com centro em \mathbf{a} . Em \mathbf{R}^2 é um disco circular, e em \mathbf{R}^3 é uma esfera de centro \mathbf{a} e raio r .

DEFINIÇÃO DE UM PONTO INTERIOR. *Seja S um subconjunto de \mathbf{R}^n e admita-se que $\mathbf{a} \in S$. Então \mathbf{a} diz-se um ponto interior de S se existir uma n -bola aberta com centro em \mathbf{a} , cujos pontos pertencem todos a S .*

Por outras palavras, todo o ponto interior \mathbf{a} de S pode ser envolvido por uma n -bola $B(\mathbf{a})$ tal que $B(\mathbf{a}) \subseteq S$. O conjunto de todos os pontos interiores de S diz-se o *interior* de S e representa-se por $\text{int } S$. Um conjunto aberto contendo um ponto \mathbf{a} designa-se muitas vezes por *vizinhança de \mathbf{a}* .

DEFINIÇÃO DE UM CONJUNTO ABERTO. *Um conjunto S de \mathbf{R}^n diz-se aberto se todos os seus pontos são pontos interiores. Por outras palavras, S é aberto se e só se $S = \text{int } S$.*

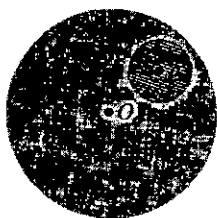
EXEMPLO. Em \mathbf{R}^1 o tipo mais simples de conjunto aberto é um intervalo aberto. A união de dois ou mais intervalos abertos é também aberto. Um intervalo fechado $[a, b]$ não é um conjunto aberto porque nenhum dos pontos extremos do intervalo pode encerrar-se numa 1-bola situada inteiramente no intervalo dado.

A 2-bola $S = B(O, 1)$ desenhada na figura 8.1 é um exemplo de um conjunto aberto de \mathbf{R}^2 . Todo o ponto a de S é o centro de um disco situado completamente em S . Para alguns pontos o raio deste disco é muito pequeno.

Em \mathbf{R}^2 podem construir-se alguns conjuntos abertos tomando o produto cartesiano de conjuntos abertos de \mathbf{R}^1 . Se A_1 e A_2 são subconjuntos de \mathbf{R}^1 , o seu produto cartesiano $A_1 \times A_2$ é o conjunto definido em \mathbf{R}^2 por

$$A_1 \times A_2 = \{(a_1, a_2) \mid a_1 \in A_1 \text{ e } a_2 \in A_2\}.$$

Na figura 8.2 apresenta-se um exemplo. Os conjuntos A_1 e A_2 são intervalos e $A_1 \times A_2$ é um rectângulo.



Disco circular

FIG. 8.1. O disco $B(O, 1)$ é um conjunto aberto em \mathbf{R}^2 .

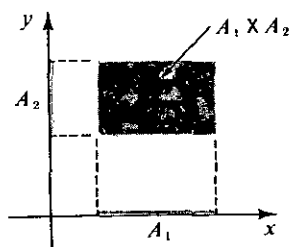


FIG. 8.2. O produto cartesiano de dois intervalos abertos é um rectângulo aberto.

Se A_1 e A_2 são subconjuntos abertos de \mathbf{R}^1 , então $A_1 \times A_2$ será um subconjunto aberto de \mathbf{R}^2 . Para demonstrar isto, escolhamos um ponto qualquer $a = (a_1, a_2)$ em $A_1 \times A_2$. Devemos mostrar que a é um ponto interior de $A_1 \times A_2$. Uma vez que A_1 e A_2 são abertos em \mathbf{R}^1 existe uma 1-bola $B(a_1; r_1)$ em A_1 e uma 1-bola $B(a_2; r_2)$ em A_2 . Seja $r = \min\{r_1, r_2\}$. Pode facilmente provar-se que a 2-bola $B(a; r) \subseteq A_1 \times A_2$. Com efeito, se $x = (x_1, x_2)$ é um qualquer ponto de $B(a; r)$ então $\|x - a\| < r$, pelo que $|x_1 - a_1| < r_1$ e $|x_2 - a_2| < r_2$. Logo $x_1 \in B(a_1; r_1)$ e $x_2 \in B(a_2; r_2)$. Deste modo $x_1 \in A_1$ e $x_2 \in A_2$, pelo que $(x_1, x_2) \in A_1 \times A_2$. Isto prova que cada ponto de $B(a; r)$ está em $A_1 \times A_2$. Por conseguinte todo o ponto de $A_1 \times A_2$ é um ponto interior, pelo que $A_1 \times A_2$ é aberto.

O leitor deverá provar que um subconjunto aberto de \mathbf{R}^1 não é mais um conjunto aberto quando considerado como um subconjunto de \mathbf{R}^2 , devido a que um subconjunto de \mathbf{R}^1 não pode conter uma 2-bola.

DEFINIÇÕES DE EXTERIOR E FRONTEIRA. Um ponto x diz-se ser exterior a um conjunto S de \mathbf{R}^n se existir uma n -bola $B(x)$ que não contenha pontos de S . O conjunto de todos os pontos de \mathbf{R}^n exteriores a S chama-se o exterior de S e representa-se por $\text{ext } S$. Um ponto que não é nem ponto interior nem ponto exterior de S chama-se um ponto fronteiro de S . O conjunto de todos pontos fronteiros de S formam a fronteira de S a qual se representa por ∂S .

Estes conceitos estão representados na figura 8.1. O exterior de S é o conjunto de todos x tais que $\|x\| > 1$. A fronteira de S consiste de todos x tais que $\|x\| = 1$.

8.3. Exercícios

1. Sejam f um campo escalar definido num conjunto S e c um número real dado. O conjunto de todos os pontos x de S tais que $f(x) = c$ chama-se um *conjunto de nível* de f . (Problemas de natureza geométrica e física tratando com conjuntos de nível serão analisados neste capítulo.) Para cada um dos seguintes campos escalares, S é todo o espaço \mathbf{R}^n . Fazer o desenho relativo aos conjuntos de nível correspondendo aos valores de c dados.

- | | |
|--|---|
| (a) $f(x, y) = x^2 + y^2$, | $c = 0, 1, 4, 9$. |
| (b) $f(x, y) = e^{xy}$, | $c = e^{-2}, e^{-1}, 1, e, e^2, e^3$. |
| (c) $f(x, y) = \cos(x + y)$, | $c = -1, 0, \frac{1}{2}, \frac{1}{2}\sqrt{2}, 1$. |
| (d) $f(x, y, z) = x + y + z$, | $c = -1, 0, 1$. |
| (e) $f(x, y, z) = x^2 + 2y^2 + 3z^2$, | $c = 0, 6, 12$. |
| (f) $f(x, y, z) = \sin(x^2 + y^2 + z^2)$, | $c = -1, -\frac{1}{2}, 0, \frac{1}{2}\sqrt{2}, 1$. |

2. Em cada uma das alíneas seguintes S é o conjunto de todos os pontos (x, y) do plano que verificam as desigualdades dadas. Fazer um gráfico representando o conjunto S e explicar, geometricamente, se S é ou não aberto. Indicar no gráfico a fronteira de S .

- | | |
|-----------------------------------|---|
| (a) $x^2 + y^2 < 1$. | (h) $1 \leq x \leq 2$ e $3 < y < 4$. |
| (b) $3x^2 + 2y^2 < 6$. | (i) $1 < x < 2$ e $y > 0$. |
| (c) $ x < 1$ e $ y < 1$. | (j) $x \geq y$. |
| (d) $x \geq 0$ e $y > 0$. | (k) $x > y$. |
| (e) $ x \leq 1$ e $ y \leq 1$. | (l) $y > x^2$ e $ x < 2$. |
| (f) $x > 0$ e $y < 0$. | (m) $(x^2 + y^2 - 1)(4 - x^2 - y^2) > 0$. |
| (g) $xy < 1$. | (n) $(2x - x^2 - y^2)(x^2 + y^2 - x) > 0$. |

3. Em cada uma das alíneas seguintes S é o conjunto de todos os pontos (x, y, z) do espaço tridimensional satisfazendo às desigualdades dadas; determinar se S é ou não aberto.

- (a) $z^2 - x^2 - y^2 - 1 > 0$.
 (b) $|x| < 1, |y| < 1$, e $|z| < 1$.
 (c) $x + y + z < 1$.
 (d) $|x| \leq 1, |y| < 1$, e $|z| < 1$.
 (e) $x + y + z < 1$ e $x > 0, y > 0, z > 0$.
 (f) $x^2 + 4y^2 + 4z^2 - 2x + 16y + 40z + 113 < 0$.

4. (a) Se A é um conjunto no espaço n -dimensional e se $x \in A$, mostrar que o conjunto $A - \{x\}$, obtido pela supressão do ponto x em A , é aberto.

- (b) Se A é um intervalo aberto no eixo real e B é um subintervalo fechado de A , mostrar que $A - B$ é aberto.†
- (c) Se A e B são intervalos abertos do eixo real, mostrar que $A \cup B$ e $A \cap B$ são abertos..
- (d) Se A é um intervalo fechado do eixo real, mostrar que o seu complemento (relativamente a todo o eixo real) é aberto.
5. Provar as seguintes propriedades de conjuntos abertos em \mathbb{R}^n .
- (a) O conjunto vazio \emptyset é aberto.
- (b) \mathbb{R}^n é aberto.
- (c) A união de uma família qualquer de conjuntos abertos é um aberto.
- (d) A intersecção de uma colecção finita de conjuntos abertos é um aberto.
- (e) Dar um exemplo para mostrar que a intersecção de uma colecção infinita de conjuntos abertos não é necessariamente um aberto.

Conjuntos fechados. Um conjunto S em \mathbb{R}^n diz-se *fechado* se o seu complemento $\mathbb{R}^n - S$ é aberto. Os três exercícios que se seguem referem-se a propriedades de conjuntos fechados.

6. Em cada uma das alíneas seguintes, seja S o conjunto de todos os pontos (x, y) em \mathbb{R}^2 satisfazendo às condições dadas. Fazer um gráfico mostrando o conjunto S e explicar geometricamente se S é aberto, fechado, simultaneamente aberto e fechado, ou nem aberto nem fechado.
- (a) $x^2 + y^2 \geq 0$. (g) $1 \leq x \leq 2, 3 \leq y \leq 4$.
- (b) $x^2 + y^2 < 0$. (h) $1 \leq x \leq 2, 3 \leq y < 4$.
- (c) $x^2 + y^2 \leq 1$. (i) $y = x^2$.
- (d) $1 < x^2 + y^2 < 2$. (j) $y \geq x^2$.
- (e) $1 \leq x^2 + y^2 \leq 2$. (k) $y \geq x^2$ and $|x| < 2$.
- (f) $1 < x^2 + y^2 \leq 2$. (l) $y \geq x^2$ and $|x| \leq 2$.
7. (a) Se A é um conjunto fechado no espaço n -dimensional e x é um ponto não pertencente a A , provar que $A \cup \{x\}$ é também fechado.
- (b) Provar que o intervalo fechado $[a, b]$ no eixo real é um conjunto fechado.
- (c) Se A e B são intervalos fechados do eixo real, mostrar que $A \cup B$ e $A \cap B$ são fechados.
8. Provar as seguintes propriedades dos conjuntos fechados em \mathbb{R}^n . Podem utilizar-se os resultados do Exercício 5.
- (a) O conjunto vazio \emptyset é fechado.
- (b) \mathbb{R}^n é fechado.
- (c) A intersecção de uma colecção qualquer de conjuntos fechados é um fechado.
- (d) A união de um número finito de conjuntos fechados é um conjunto fechado.
- (e) Dar um exemplo para mostrar que a união de uma colecção infinita de conjuntos fechados não é necessariamente fechada.
9. Seja S um subconjunto de \mathbb{R}^n .
- (a) Provar que ambos $\text{int } S$ e $\text{ext } S$ são conjuntos abertos.
- (b) Provar que $\mathbb{R}^n = (\text{int } S) \cup (\text{ext } S) \cup \partial S$, uma união de conjuntos disjuntos, e utilizar este facto para provar que a fronteira ∂S é sempre um conjunto fechado.
10. Dado um conjunto S em \mathbb{R}^n e um ponto x gozando da propriedade de que toda a bola $B(x)$ contém pontos interiores e exteriores a S simultaneamente, provar que x é um ponto

† Se A e B são conjuntos, a diferença $A - B$ (chamado o *complemento de B relativamente a A*) é o conjunto de todos os elementos de A que não pertencem a B .

fronteiro de S . Será a inversa verdadeira, isto é, cada ponto fronteiro de S gozará necessariamente desta propriedade?

11. Seja S um subconjunto de \mathbb{R}^n . Provar que $\text{ext } S = \text{int}(\mathbb{R}^n - S)$.

12. Provar que um conjunto S em \mathbb{R}^n é fechado se e só se $S = (\text{int } S) \cup \partial S$.

8.4. Limites e continuidade

Os conceitos de limite e continuidade são facilmente generalizáveis a campos escalares e vectoriais. Vamos formular as definições para campos vectoriais; elas serão também aplicáveis a campos escalares.

Consideremos uma função $f: S \rightarrow \mathbb{R}^m$, onde S é um subconjunto de \mathbb{R}^n . Se $a \in \mathbb{R}^n$ e $b \in \mathbb{R}^m$ escrevemos

$$(8.1) \quad \lim_{x \rightarrow a} f(x) = b \quad (\text{ou, } f(x) \rightarrow b \text{ quando } x \rightarrow a)$$

para

$$(8.2) \quad \lim_{\|x-a\| \rightarrow 0} \|f(x) - b\| = 0.$$

O símbolo limite na equação (8.2) é o limite usual do cálculo elementar. Nesta definição não se exige que f seja definida no próprio ponto a .

Se escrevermos $h = x - a$, (8.2) vem

$$\lim_{\|h\| \rightarrow 0} \|f(a+h) - b\| = 0.$$

Para pontos de \mathbb{R}^2 escrevemos (x, y) em vez de x e (a, b) em vez de a e exprimimos a relação limite (8.1) do modo seguinte:

$$\lim_{(x,y) \rightarrow (a,b)} f(x, y) = b.$$

Para ponto de \mathbb{R}^3 pomos $x = (x, y, z)$ e $a = (a, b, c)$ e escrevemos

$$\lim_{(x,y,z) \rightarrow (a,b,c)} f(x, y, z) = b.$$

Uma função f diz-se *contínua* em a se f é definida em a e se

$$\lim_{x \rightarrow a} f(x) = f(a).$$

Dizemos que f é *contínua num conjunto* S se f for contínua em cada ponto de S .

Porque estas definições são extensões directas das dadas para o caso unidimensional, não surpreenderá que muitas das propriedades dos limites e da continuidade possam igualmente generalizar-se. Por exemplo, os teoremas relativos a limites e continuidade de somas, produtos e cocientes são também válidos para campos escalares. Para campos vectoriais não se definem os cocientes, mas podem estabelecer-se os teoremas

seguintes respeitantes a somas, multiplicação por escalares, produtos internos e normas.

TEOREMA 8.1. Se $\lim_{x \rightarrow a} f(x) = b$ e $\lim_{x \rightarrow a} g(x) = c$, então tem-se

- (a) $\lim_{x \rightarrow a} [f(x) + g(x)] = b + c$.
- (b) $\lim_{x \rightarrow a} \lambda f(x) = \lambda b$ para todo escalar λ .
- (c) $\lim_{x \rightarrow a} f(x) \cdot g(x) = b \cdot c$.
- (d) $\lim_{x \rightarrow a} \|f(x)\| = \|b\|$.

Demonstração. Demonstraremos unicamente (c) e (d); as demonstrações de (a) e (b) são deixadas ao leitor como exercício.

Para demonstrar (c) escrevemos

$$f(x) \cdot g(x) - b \cdot c = [f(x) - b] \cdot [g(x) - c] + b \cdot [g(x) - c] + c \cdot [f(x) - b].$$

Apliquemos agora a desigualdade triangular e a desigualdade de Cauchy-Schwarz para obtermos

$$0 \leq \|f(x) \cdot g(x) - b \cdot c\| \leq \|f(x) - b\| \|g(x) - c\| + \|b\| \|g(x) - c\| + \|c\| \|f(x) - b\|.$$

Porque $\|f(x) - b\| \rightarrow 0$ e $\|g(x) - c\| \rightarrow 0$ quando $x \rightarrow a$, isto prova que $\|f(x) \cdot g(x) - b \cdot c\| \rightarrow 0$ quando $x \rightarrow a$, o que demonstra (c).

Tomando $f(x) = g(x)$ na alínea (c) encontramos

$$\lim_{x \rightarrow a} \|f(x)\|^2 = \|b\|^2,$$

donde se obtém (d).

EXEMPLO 1. *Continuidade e componentes de um campo vectorial.* Se um campo vectorial f tem valores em \mathbb{R}^m , cada um dos valores da função $f(x)$ tem m componentes e podemos escrever

$$f(x) = (f_1(x), \dots, f_m(x)).$$

Os m campos escalares f_1, \dots, f_m dizem-se *componentes* do campo vectorial f . Provaremos que f é contínua num ponto se, e só se, cada componente f_k é contínua naquele ponto.

Designemos por e_k o vector unitário coordenado de ordem k (todas as suas componentes são zeros excepto a de ordem k que é igual a 1). Então $f_k(x)$ é dada pelo produto escalar

$$f_k(x) = f(x) \cdot e_k.$$

Deste modo a alínea (c) do teorema 8.1 mostra que cada ponto de continuidade de f é também um ponto de continuidade de f_k . Além disso, uma vez que se tem

$$f(x) = \sum_{k=1}^m f_k(x)e_k,$$

a aplicação repetidas das alíneas (a) e (b) do teorema 8.1 mostra que um ponto de continuidade de todas as m componentes f_1, \dots, f_m é também um ponto de continuidade de f .

EXEMPLO 2. *Continuidade da função identidade.* A função identidade, $f(x) = x$, é contínua em todo \mathbf{R}^n . Consequentemente, as suas componentes são também contínuas em todo \mathbf{R}^n . Elas constituem n campos escalares definidos por

$$f_1(x) = x_1, f_2(x) = x_2, \dots, f_n(x) = x_n.$$

EXEMPLO 3. *Continuidade das transformações lineares.* Seja $f: \mathbf{R}^n \rightarrow \mathbf{R}^m$ uma transformação linear. Vamos demonstrar que f é contínua em cada ponto a de \mathbf{R}^n . Devido à linearidade temos

$$f(a + h) = f(a) + f(h).$$

Basta portanto demonstrar que $f(h) \rightarrow 0$ quando $h \rightarrow 0$. Escrevendo h em função das suas componentes temos $h = h_1 e_1 + \dots + h_n e_n$. Utilizando de novo a linearidade temos $f(h) = h_1 f(e_1) + \dots + h_n f(e_n)$. Isto mostra que $f(h) \rightarrow 0$ quando $h \rightarrow 0$.

EXEMPLO 4. *Continuidade dos polinômios de n variáveis.* Um campo escalar P , definido em \mathbf{R}^n por uma fórmula da forma

$$P(x) = \sum_{k_1=0}^{p_1} \dots \sum_{k_n=0}^{p_n} c_{k_1 \dots k_n} x_1^{k_1} \dots x_n^{k_n}$$

chama-se polinômio de n variáveis x_1, \dots, x_n . Um polinômio é contínuo em todo \mathbf{R}^n devido a que é uma soma finita de produtos de campos escalares contínuos em todo \mathbf{R}^n . Por exemplo, um polinômio de duas variáveis x e y , dado por

$$P(x, y) = \sum_{i=0}^p \sum_{j=0}^q c_{ij} x^i y^j$$

é contínuo em todo o ponto (x, y) de \mathbf{R}^2 .

EXEMPLO 5. *Continuidade de funções racionais.* Um campo escalar f definido por $f(x) = P(x)/Q(x)$, onde P e Q são polinómios nas componentes de x , diz-se uma função racional. Uma tal função é contínua em cada ponto em que $Q(x) \neq 0$.

Mais exemplos de funções contínuas podem ser construídos com auxílio do teorema que se segue, o qual está relacionado com a continuidade de funções compostas.

TEOREMA 8.2. *Sejam f e g funções tais que a função composta $f \circ g$ esteja definida em a , sendo*

$$(f \circ g)(x) = f[g(x)].$$

Se g é contínua em a e se f é contínua em $g(a)$, então a função composta $f \circ g$ é contínua em a .

Demonstração. Sejam $y = f(x)$ e $b = g(a)$. Temos pois:

$$f[g(x)] - f[g(a)] = f(y) - f(b).$$

Por hipótese $y \rightarrow b$ quando $x \rightarrow a$, pelo que se terá

$$\lim_{\|x-a\| \rightarrow 0} \|f[g(x)] - f[g(a)]\| = \lim_{\|y-b\| \rightarrow 0} \|f(y) - f(b)\| = 0.$$

Portanto $\lim_{x \rightarrow a} f[g(x)] = f[g(a)]$, pelo que $f \circ g$ é contínua em a .

EXEMPLO 6. O teorema precedente implica a continuidade dos campos escalares h , onde $h(x, y)$ é dado por fórmulas tais como

$$\sin(x^2y), \quad \log(x^2 + y^2), \quad \frac{e^{x+y}}{x+y}, \quad \log[\cos(x^2 + y^2)].$$

Nestes exemplos as funções são contínuas em todos os pontos em que estão definidas. A primeira é contínua em todos os pontos do plano, e a segunda em todos os pontos excepto a origem. A terceira é contínua em todos os pontos (x, y) para os quais $x + y \neq 0$, e a quarta em todos os pontos para os quais $x^2 + y^2$ não seja um múltiplo ímpar de $\pi/2$. [O conjunto de (x, y) tais que $x^2 + y^2 = n\pi/2$, $n = 1, 3, 5, \dots$, é uma família de circunferências centradas na origem]. Estes exemplos mostram que as descontinuidades de uma função de duas variáveis podem ser pontos isolados, curvas, ou famílias de curvas.

EXEMPLO 7. Uma função de duas variáveis pode ser contínua a respeito de cada uma das variáveis separadamente e ser descontínua considerada como função das duas variáveis em conjunto. Este facto é ilustrado pelo exemplo seguinte

$$f(x, y) = \frac{xy}{x^2 + y^2} \quad \text{se } (x, y) \neq (0, 0), \quad f(0, 0) = 0.$$

Para pontos (x, y) do eixo OX temos $y = 0$ e $f(x, y) = f(x, 0) = 0$, pelo que a função tem o valor constante 0 ao longo de todo o eixo OX . Deste modo, se fizermos $y = 0$ e considerarmos f unicamente uma função de x , f é contínua em $x = 0$. Analogamente, f torna o valor constante 0 em todos os pontos do eixo OY , pelo que se fizermos $x = 0$ e considerarmos f unicamente como uma função de y , f é contínua em $y = 0$. Porém considerada como função de duas variáveis, f não é contínua na origem. Com efeito, em cada ponto da recta $y = x$ (excepto na origem) a função toma o valor constante $\frac{1}{2}$, pois $f(x, x) = x^2/(2x^2) = \frac{1}{2}$; visto existirem sobre esta recta pontos tão próximos da origem quanto se queira e porque $f(0, 0) \neq \frac{1}{2}$, a função não é contínua em $(0, 0)$.

8.5. Exercícios

Os exercícios apresentados nesta Secção referem-se a limites e continuidade de campos escalares definidas em subconjuntos do plano.

1. Em cada um dos exemplos seguintes define-se um campo escalar f mediante a equação dada para todos os pontos (x, y) do plano definidos pela expressão do segundo membro. Em cada exemplo determinar o conjunto de pontos (x, y) para os quais f é contínua

$$(a) f(x, y) = x^4 + y^4 - 4x^2y^2.$$

$$(f) f(x, y) = \arcsen \frac{x}{\sqrt{x^2 + y^2}}.$$

$$(b) f(x, y) = \log(x^2 + y^2).$$

$$(g) f(x, y) = \arctan \frac{x + y}{1 - xy}.$$

$$(c) f(x, y) = \frac{1}{y} \cos x^2.$$

$$(h) f(x, y) = \frac{x}{\sqrt{x^2 + y^2}}.$$

$$(d) f(x, y) = \tan \frac{x^2}{y}.$$

$$(i) f(x, y) = x^{(y^2)}.$$

$$(e) f(x, y) = \arctan \frac{y}{x}.$$

$$(j) f(x, y) = \arccos \sqrt{x/y}.$$

2. Se $\lim_{(x, y) \rightarrow (a, b)} f(x, y) = L$, e se existirem os limites unidimensionais

$$\lim_{x \rightarrow a} f(x, y) \quad \text{e} \quad \lim_{y \rightarrow b} f(x, y)$$

provar que

$$\lim_{x \rightarrow a} [\lim_{y \rightarrow b} f(x, y)] = \lim_{y \rightarrow b} [\lim_{x \rightarrow a} f(x, y)] = L.$$

Os dois limites desta igualdade chamam-se limites *iterados*; o exercício mostra que a existência do limite bidimensional e dos dois limites unidimensionais implicam a existência e

igualdade dos dois limites iterados. (A inversa não é sempre verdadeira. No exercício 4 apresenta-se um contra exemplo).

3. Seja $f(x, y) = (x - y)/(x + y)$ se $x + y \neq 0$. Provar que

$$\lim_{x \rightarrow 0} [\lim_{y \rightarrow 0} f(x, y)] = 1 \quad \text{mas que} \quad \lim_{y \rightarrow 0} [\lim_{x \rightarrow 0} f(x, y)] = -1.$$

Utilizar este resultado e o do Exercício 2 para provar que $f(x, y)$ não tende para um limite quando $(x, y) \rightarrow (0, 0)$.

4. Seja

$$f(x, y) = \frac{x^2 y^2}{x^2 y^2 + (x - y)^2} \quad \text{sempre que} \quad x^2 y^2 + (x - y)^2 \neq 0.$$

Demonstrar que

$$\lim_{x \rightarrow 0} [\lim_{y \rightarrow 0} f(x, y)] = \lim_{y \rightarrow 0} [\lim_{x \rightarrow 0} f(x, y)] = 0$$

mas que $f(x, y)$ não tende para um limite quando $(x, y) \rightarrow (0, 0)$. [Sugestão: Examinar f sobre a recta $y = x$]. Serve este exercício para mostrar que a inversa do Exercício 2 nem sempre é verdadeira.

5. Seja

$$f(x, y) = \begin{cases} x \operatorname{sen} \frac{1}{y} & \text{se } y \neq 0, \\ 0 & \text{se } y = 0. \end{cases}$$

Mostrar que $f(x, y) \rightarrow 0$ quando $(x, y) \rightarrow (0, 0)$, mas que

$$\lim_{y \rightarrow 0} [\lim_{x \rightarrow 0} f(x, y)] \neq \lim_{x \rightarrow 0} [\lim_{y \rightarrow 0} f(x, y)].$$

Explicar porque razão isto não contradiz o Exercício 2.

- Se $(x, y) \neq (0, 0)$, seja $f(x, y) = (x^2 - y^2)/(x^2 + y^2)$. Determinar o limite de $f(x, y)$ quando $(x, y) \rightarrow (0, 0)$ ao longo de recta $y = mx$. É possível definir $f(0, 0)$ de modo que f seja contínua em $(0, 0)$?
- Seja $f(x, y) = 0$ se $y \leq 0$ ou se $y \geq x^2$ e $f(x, y) = 1$ se $0 < y < x^2$. Provar que $f(x, y) \rightarrow 0$ quando $(x, y) \rightarrow (0, 0)$ ao longo de qualquer recta passando pela origem. Determinar uma curva passando pela origem ao longo da qual (excepto na origem) $f(x, y)$ tome o valor constante 1. Será f contínua na origem?
- Se $f(x, y) = |\operatorname{sen}(x^2 + y^2)|/(x^2 + y^2)$ quando $(x, y) \neq (0, 0)$ como deve ser definida $f(0, 0)$ para que f seja contínua na origem?
- Seja f um campo escalar contínuo num ponto interior a de um conjunto S de \mathbb{R}^n . Se $f(a) \neq 0$, provar que existe uma n -bola $B(a)$ na qual f tem o mesmo sinal que $f(a)$.

8.6. A derivada de um campo escalar relativamente a um vector

Esta Secção é dedicada às derivadas de campos escalares. Derivadas de campos vectoriais serão estudados na Secção 8.18.

Seja f um campo escalar definido num conjunto S de \mathbb{R}^n e seja a um ponto interior a S . Pretendemos estudar como varia o campo quando se passa de a para um

ponto próximo. Por exemplo, admitamos que $f(a)$ é a temperatura num ponto a de uma sala aquecida e com uma janela aberta. Se nos deslocarmos na direcção da janela a temperatura diminuirá, mas se nos movermos em sentido contrário a temperatura aumentará. Em geral a maneira segundo a qual um campo varia dependerá da direcção em que nos movamos a partir de a .

Supongamos que especificamos esta direcção por outro vector y , isto é, suponhamos que nos deslocamos de a para o ponto $a + y$ ao longo do segmento de recta unindo a e $a + y$. Cada ponto deste segmento é da forma $a + hy$, onde h é um número real. Na figura 8.3 apresenta-se um exemplo. A distância de a até $a + hy$ é $\|hy\| = |h| \|y\|$.

Visto que a é interior a S , existe uma n -bola $B(a; r)$ situado inteiramente em S . Se h for escolhido de modo que $|h| \|y\| < r$, o segmento de a até $a + hy$ estará em S .

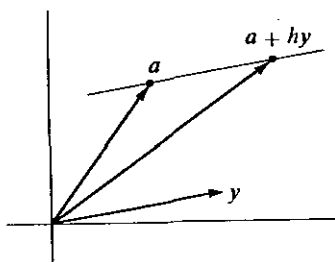


FIG. 8.3. O ponto $a + hy$ pertence à recta passando por a paralela a y .

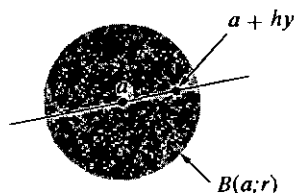


FIG. 8.4. O ponto $a + hy$ está na n -bola $B(a; r)$ se $\|hy\| < r$.

(Ver figura 8.4) Consideremos $h \neq 0$, mas suficientemente pequeno para garantir que $a + hy \in S$ e formemos a razão incremental

$$(8.3) \quad \frac{f(a + hy) - f(a)}{h}$$

O numerador deste cociente diz-nos como varia a função quando se passa de a a $a + hy$. O cociente define a *variação média* de f ao longo do segmento unindo a a $a + hy$. Interessa-nos conhecer o comportamento deste cociente quando $h \rightarrow 0$. Somos assim conduzidos a definição seguinte:

DEFINIÇÃO DE DERIVADA DE UM CAMPO ESCALAR RELATIVAMENTE A UM VECTOR. Dado um campo escalar $f: S \rightarrow \mathbb{R}$, onde $S \subseteq \mathbb{R}^n$, sejam a um ponto interior de S e y um ponto arbitrário de \mathbb{R}^n . A derivada de f em a relativamente a y representa-se pelo símbolo $f'(a; y)$ e define-se

$$(8.4) \quad f'(a; y) = \lim_{h \rightarrow 0} \frac{f(a + hy) - f(a)}{h}$$

quando o limite do segundo membro existe.

EXEMPLO 1. Se $y = 0$, a razão incremental (8.3) é 0 para todo $h \neq 0$, pelo que $f'(a; 0)$ existe sempre e é igual a 0.

EXEMPLO 2. *Derivada de uma transformação linear.* Se $f: S \rightarrow \mathbb{R}$ é linear então $f(a + hy) = f(a) + hf(y)$ e a razão incremental (8.3) é igual a $f(y)$ para todo $h \neq 0$. Neste caso, $f'(a; y)$ existe sempre e é dada por

$$f'(a; y) = f(y)$$

para todo a de S e todo y de \mathbb{R}^n . Quer dizer, a derivada de uma transformação linear relativamente a y é igual ao valor da função em y .

Para estudar o comportamento de f sobre a recta que passa por a e $a + y$ com $y \neq 0$, introduzimos a função

$$g(t) = f(a + ty).$$

O teorema que se apresenta a seguir relativa às derivadas $g'(t)$ e $f'(a + ty; y)$.

TEOREMA 8.3. *Seja $g(t) = f(a + ty)$. Se uma das derivadas $g'(t)$ ou $f'(a + ty; y)$ existir, então a outra também existe e as duas são iguais*

$$(8.5) \quad g'(t) = f'(a + ty; y).$$

Em particular, quando $t = 0$ temos $g'(0) = f'(a; y)$.

Demonstração. Construindo a razão incremental para g , temos

$$\frac{g(t+h) - g(t)}{h} = \frac{f(a + ty + hy) - f(a + ty)}{h}.$$

Fazendo $h \rightarrow 0$ obtemos (8.5).

EXEMPLO 3. Calcular $f'(a; y)$ se $f(x) = \|x\|^2$ para todo x em \mathbb{R}^n .

Resolução. Fazemos $g(t) = f(a + ty) = (a + ty) \cdot (a + ty) = a \cdot a + 2ta \cdot y + t^2 y \cdot y$. Por conseguinte $g'(t) = 2a \cdot y + 2ty \cdot y$, de maneira que $g'(0) = f'(a; y) = 2a \cdot y$.

Um corolário simples do teorema 8.3 é o teorema da média para campos escalares.

TEOREMA 8.4. TEOREMA DA MÉDIA PARA AS DERIVADAS DE CAMPOS ESCALARES. *Se existir a derivada $f'(a + ty; y)$ para todo t no intervalo $0 \leq t \leq 1$, então para algum θ no intervalo aberto $0 < \theta < 1$ tem-se*

$$f(a + y) - f(a) = f'(z; y), \quad \text{onde } z = a + \theta y.$$

Demonstração. Seja $g(t) = f(a + ty)$. Aplicando o teorema da média a g no intervalo $[0, 1]$ tem-se

$$g(1) - g(0) = g'(\theta), \quad \text{onde } 0 < \theta < 1.$$

Porque $g(1) - g(0) = f(a + y) - f(a)$ e $g'(\theta) = f'(a + \theta y; y)$ fica o teorema demonstrado.

8.7. Derivadas direccionais e derivadas parciais

No caso particular em que y é um vector unitário isto é, quando $\|y\| = 1$, a distância de a a $a + hy$ é $|h|$. Neste caso a razão incremental (8.3) representa a variação média de f por unidade de distância ao longo do segmento unindo a a $a + hy$; a derivada $f'(a; y)$ chama-se *derivada direcciona*l.

DEFINIÇÃO DAS DERIVADAS DIRECCIONAL E PARCIAL. Se y é um vector unitário, a derivada $f'(a; y)$ chama-se a derivada direcciona

$$D_k f(a) = f'(a; e_k).$$

Para a derivada parcial $D_k f(a)$ usam-se igualmente as notações:

$$D_k f(a_1, \dots, a_n), \quad \frac{\partial f}{\partial x_k}(a_1, \dots, a_n), \quad \text{e} \quad f'_{x_k}(a_1, \dots, a_n).$$

Por vezes a derivada f'_{x_k} escreve-se sem ápice, f_{x_k} , ou ainda mais simplesmente f_k .

Em \mathbb{R}^2 os vectores unitários coordenados representam-se habitualmente por i e j . Se $a = (a, b)$, as derivadas parciais $f'(a; i)$ e $f'(a; j)$ escrevem-se também

$$\frac{\partial f}{\partial x}(a, b) \quad \text{e} \quad \frac{\partial f}{\partial y}(a, b),$$

respectivamente. Em \mathbb{R}^3 , se $a = (a, b, c)$ as derivadas parciais $D_1 f(a)$, $D_2 f(a)$ e $D_3 f(a)$ representam-se também por

$$\frac{\partial f}{\partial x}(a, b, c), \quad \frac{\partial f}{\partial y}(a, b, c), \quad \text{e} \quad \frac{\partial f}{\partial z}(a, b, c).$$

8.8 Derivadas parciais de ordem superior

A derivação parcial origina, a partir de um dado campo f , novos campos escalares $D_1 f, \dots, D_n f$. As derivadas parciais de $D_1 f, \dots, D_n f$ são as *derivadas parciais de segunda ordem de f* . Para funções de duas variáveis existem quatro derivadas parciais de segunda ordem, que se escrevem:

$$D_1(D_1 f) = \frac{\partial^2 f}{\partial x^2}, \quad D_1(D_2 f) = \frac{\partial^2 f}{\partial x \partial y}, \quad D_2(D_1 f) = \frac{\partial^2 f}{\partial y \partial x}, \quad D_2(D_2 f) = \frac{\partial^2 f}{\partial y^2}.$$

Note-se que $D_1(D_2 f)$ representa a derivada parcial de $D_2 f$ relativamente à primeira variável. Algumas vezes usamos a notação $D_{ij} f$ para a derivada parcial de segunda ordem $D_i(D_j f)$. Por exemplo, $D_{12} f = D_1(D_2 f)$. Na notação ∂ a ordem de derivação exprime-se como se indica:

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial y} \right).$$

Esta derivada pode ou não ser igual à outra derivada parcial mista

$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial x} \right).$$

Na Secção 8.23 provaremos que as duas derivadas parciais mistas acabadas de referir são iguais em cada ponto se uma delas for contínua numa vizinhança do ponto. Ainda na Secção 8.23 se apresentará um exemplo no qual $D_1(D_2 f) \neq D_2(D_1 f)$ num ponto.

8.9 Exercícios

- Um campo escalar f é definido em \mathbb{R}^n por $f(x) = a \cdot x$, onde a é um vector constante. Calcular $f'(x, y)$ para x e y arbitrários.
- (a) Resolver o Exercício 1 quando $f(x) = \|x\|^4$.
(b) Tomar $n = 2$ em (a) e determinar todos os pontos (x, y) para os quais $f'(2i + 3j; xi + yj) = 6$.
(c) Tomar $n = 3$ em (a) e determinar todos os pontos (x, y, z) para os quais $f'(i + 2j + 3k; xi + yj + zk) = 0$.
- Seja $T: \mathbb{R}^n \rightarrow \mathbb{R}^n$ uma transformação linear dada. Calcular a derivada $f'(x; y)$ para o campo escalar definido em \mathbb{R}^n pela equação $f(x) = x \cdot T(x)$.

Em cada um dos Exercícios 4 a 9, calcular todas as derivadas de primeira ordem do campo escalar dado. Nos Exercícios 8 e 9 os campos estão definidos em \mathbb{R}^n .

- $f(x, y) = x^2 + y^2 \sin(xy)$.
- $f(x, y) = \frac{x+y}{x-y}, \quad x \neq y$.
- $f(x, y) = \sqrt{x^2 + y^2}$.
- $f(x) = a \cdot x, \quad a$ fixo.
- $f(x, y) = \frac{x}{\sqrt{x^2 + y^2}}, \quad (x, y) \neq (0, 0)$.
- $f(x) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j, \quad a_{ij} = a_{ji}$.

Em cada um dos Exercícios 10 a 17, calcular todas as derivadas parciais de primeira ordem. Nos Exercícios 10, 11 e 12 verificar que as derivadas parciais $D_1(D_2 f)$ são iguais.

$$10. f(x, y) = x^4 + y^4 - 4x^2y^2.$$

$$14. f(x, y) = \arctan(y/x), \quad x \neq 0.$$

$$11. f(x, y) = \log(x^2 + y^2), \quad (x, y) \neq (0, 0). \quad 15. f(x, y) = \arctan \frac{x+y}{1-xy}, \quad xy \neq 1.$$

$$12. f(x, y) = \frac{1}{y} \cos x^2, \quad y \neq 0.$$

$$16. f(x, y) = x^{(y^2)}, \quad x > 0.$$

$$13. f(x, y) = \tan(x^2/y), \quad y \neq 0.$$

$$17. f(x, y) = \arccos \sqrt{x/y}, \quad y \neq 0.$$

18. Seja $v(r, t) = t^n e^{-r^2/(4t)}$. Determinar um valor da constante n tal que v satisfaça a seguinte equação:

$$\frac{\partial v}{\partial t} = \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial v}{\partial r} \right).$$

19. Dado $z = u(x, y)e^{ax+by}$ e $\partial^2 u / (\partial x \partial y) = 0$. Determinar valores das constantes a e b tais que

$$\frac{\partial^2 z}{\partial x \partial y} - \frac{\partial z}{\partial x} - \frac{\partial z}{\partial y} + z = 0.$$

20. (a) Supor que $f(x; y) = 0$ para todo x em alguma n -bola $B(a)$ para todo o vector y . Utilizar o teorema da média para provar que f é constante em $B(a)$.

(b) Supor que $f(x; y) = 0$ para um vector fixo y e para todo x em $B(a)$. O que pode concluir-se acerca de f neste caso?

21. Um conjunto S em \mathbb{R}^n diz-se *convexo* se para todo o par de pontos a e b em S o segmento de recta de a a b está também em S ; por outras palavras, $ta + (1-t)b \in S$, para todo t no intervalo $0 \leq t \leq 1$.

(a) Provar que toda a n -bola é convexa.

(b) Se $f(x; y) = 0$ para todo x num conjunto convexo aberto S e para todo y em \mathbb{R}^n , provar que f é constante em S .

22. (a) Demonstrar que não existe nenhum campo escalar f tal que $f'(a; y) > 0$ para um vector fixo a e todo vector não nulo y .

(b) Dar um exemplo de um campo escalar f tal que $f'(x; y) > 0$ para um vector fixo y e todo o vector x .

8.10. Derivadas direccionais e continuidade

Na teoria unidimensional, a existência de derivada de uma função f num ponto implica a continuidade nesse ponto. Tal afirmação demonstra-se facilmente pela escolha de um $h \neq 0$ e escrevendo

$$f(a+h) - f(a) = \frac{f(a+h) - f(a)}{h} \cdot h.$$

Cuando $h \rightarrow 0$ o segundo membro tende para o limite $f'(a) \cdot 0 = 0$ e por isso $f(a+h) \rightarrow f(a)$. Isto prova que a existência de $f'(a)$ implica a continuidade de f em a .

Admitamos o mesmo raciocínio aplicado a uma campo escalar qualquer. Suponhamos que a derivada $f'(a; y)$ existe para certo y . Então se $h \neq 0$ podemos escrever

$$f(a + hy) - f(a) = \frac{f(a + hy) - f(a)}{h} \cdot h.$$

Quando $h \rightarrow 0$ o segundo membro tende para o limite $f'(a; y) \cdot 0 = 0$; logo a existência de $f'(a; y)$ para um dado y implica que

$$\lim_{h \rightarrow 0} f(a + hy) = f(a)$$

para algum y . Tal significa que $f(x) \rightarrow f(a)$ quando $x \rightarrow a$ ao longo da recta passando por a e paralela a y . Se $f'(a; y)$ existe para todo o vector y , então $f(x) \rightarrow f(a)$ quando $x \rightarrow a$ ao longo de toda a recta que passe por a . Tal facto parece sugerir que f é contínua em a . Surpreendentemente, embora, esta conclusão não é necessariamente verdadeira. O exemplo que se apresenta a seguir descreve um campo escalar que admite derivada direccional em O segundo qualquer direcção mas que não é contínuo em O .

EXEMPLO. Seja f o campo escalar definido em \mathbf{R}^2 do modo seguinte:

$$f(x, y) = \frac{xy^2}{x^2 + y^4} \quad \text{se } x \neq 0, \quad f(0, y) = 0.$$

Seja $a = (0, 0)$ e $y = (a, b)$ um vector qualquer. Se $a \neq 0$ e $h \neq 0$ temos

$$\frac{f(a + hy) - f(a)}{h} = \frac{f(hy) - f(O)}{h} = \frac{f(ha, hb)}{h} = \frac{ab^2}{a^2 + h^2b^4}.$$

Fazendo $h \rightarrow 0$ encontramos $f'(O; y) = b^2/a$. Se $y = (0, b)$ encontramos, de maneira análoga, que $f'(O; y) = 0$. Portanto $f'(O; y)$ existe para todas as direcções y . Igualmente, $f(x) \rightarrow 0$ quando $x \rightarrow O$ ao longo de qualquer recta passando pela origem. Contudo, em cada ponto da parábola $x = y^2$ (excepto na origem) a função f tem o valor $\frac{1}{2}$. Visto que tais pontos existem tão próximos da origem quanto se queria e porque $f(O) = 0$, a função f não é contínua em O .

O exemplo precedente mostra que a existência de *todas* as derivadas direccionais num ponto não implica a continuidade nesse ponto. Por tal motivo, as derivadas direccionais não constituem uma extensão satisfatória do conceito unidimensional de derivada. Existe uma generalização mais adequada, a qual implica a continuidade e, ao mesmo tempo, permite-nos generalizar os principais teoremas da derivação a uma dimensão ao caso de maior número de dimensões. Esta é a chamada *diferencial total* ou simplesmente *diferencial*.

8.11. A diferencial

Lembramos que no caso unidimensional uma função f possuindo derivada em a pode ser aproximada na vizinhança de a mediante um polinómio de Taylor linear. Se $f'(a)$ existe, designemos por $E(a, h)$ a diferença

$$(8.6) \quad E(a, h) = \frac{f(a+h) - f(a)}{h} - f'(a) \quad \text{se } h \neq 0,$$

e definamos $E(a, 0) = 0$. De (8.6) obtemos a fórmula

$$f(a+h) = f(a) + f'(a)h + hE(a, h),$$

uma equação que é válida também para $h = 0$. Esta é a fórmula de Taylor de primeira ordem para aproximar $f(a+h) - f(a)$ por meio de $f'(a)h$. O erro cometido é $hE(a, h)$. De (8.6) resulta que $E(a, h) \rightarrow 0$ quando $h \rightarrow 0$. Por conseguinte o erro $hE(a, h)$ é de ordem inferior a h para valores pequenos de h .

Esta propriedade de aproximação de uma função derivável por uma função linear sugere um modo de generalização do conceito de diferenciabilidade ao caso de maior número de dimensões.

Seja $f: S \rightarrow \mathbf{R}$ um campo escalar definido num conjunto S de \mathbf{R}^n . Sejam a um ponto interior de S e $B(a; r)$ uma n -bola contida em S . Seja v um vector tal que $\|v\| < r$, de modo que $a + v \in B(a; r)$.

DEFINIÇÃO DE UM CAMPO ESCALAR DIFERENCIÁVEL. Diz-se que f é diferenciável em a se existir uma transformação

$$T_a: \mathbf{R}^n \rightarrow \mathbf{R}$$

de \mathbf{R}^n em \mathbf{R} , e uma função escalar $E(a, v)$ tal que

$$(8.7) \quad f(a+v) = f(a) + T_a(v) + \|v\| E(a, v),$$

para $\|v\| < r$, onde $E(a, v) \rightarrow 0$ quando $\|v\| \rightarrow 0$. A transformação linear T_a chama-se a diferencial de f em a .

Nota: A diferencial T_a é uma transformação linear, não um número. O valor $T_a(v)$ é um número real, definido para cada ponto v de \mathbf{R}^n . A diferencial foi introduzida por W. H. Young em 1908 e por M. Fréchet em 1911 num contexto mais geral.

A equação (8.7), que é válida para $\|v\| < r$, chama-se *fórmula de Taylor de primeira ordem* para $f(a+v)$. Dá-nos uma aproximação linear, $T_a(v)$, para a diferença $f(a+v) - f(a)$. O erro na aproximação é $\|v\| E(a, v)$, um termo que é de ordem inferior a $\|v\|$ quando $\|v\| \rightarrow 0$; isto é, $E(a, v) = O(\|v\|)$ quando $\|v\| \rightarrow 0$.

O teorema que se segue mostra que se a diferencial existe é única. Diz-se igualmente como calcular $T_a(\mathbf{y})$ para todo \mathbf{y} em \mathbb{R}^n .

TEOREMA 8.5. *Se f é diferenciável em \mathbf{a} com diferencial T_a , então a derivada $f'(\mathbf{a}; \mathbf{y})$ existe para todo \mathbf{y} em \mathbb{R}^n e tem-se*

$$(8.8) \quad T_a(\mathbf{y}) = f'(\mathbf{a}; \mathbf{y}).$$

Além disso, $f'(\mathbf{a}; \mathbf{y})$ é uma combinação linear das componentes de \mathbf{y} . Com efeito, se $\mathbf{y} = (y_1, \dots, y_n)$ tem-se

$$(8.9) \quad f'(\mathbf{a}; \mathbf{y}) = \sum_{k=1}^n D_k f(\mathbf{a}) y_k.$$

Demonstração. A equação (8.8) é trivial se $\mathbf{y} = \mathbf{0}$, visto que ambos $T_a(\mathbf{0}) = \mathbf{0}$ e $f'(\mathbf{a}; \mathbf{0}) = 0$. Por conseguinte podemos supor que $\mathbf{y} \neq \mathbf{0}$.

Visto f ser diferenciável em \mathbf{a} temos uma fórmula de Taylor

$$(8.10) \quad f(\mathbf{a} + \mathbf{v}) = f(\mathbf{a}) + T_a(\mathbf{v}) + \|\mathbf{v}\| E(\mathbf{a}, \mathbf{v})$$

para $\|\mathbf{v}\| < r$ para algum $r > 0$, onde $E(\mathbf{a}, \mathbf{v}) \rightarrow 0$ quando $\|\mathbf{v}\| \rightarrow 0$. Nesta fórmula tomemos $\mathbf{v} = h\mathbf{y}$, onde $h \neq 0$ e $|h|\|\mathbf{y}\| < r$. Então $\|\mathbf{v}\| < r$. Visto que T_a é linear, temos $T_a(\mathbf{v}) = T_a(h\mathbf{y}) = hT_a(\mathbf{y})$. Portanto (8.10) dá-nos

$$(8.11) \quad \frac{f(\mathbf{a} + h\mathbf{y}) - f(\mathbf{a})}{h} = T_a(\mathbf{y}) + \frac{|h|\|\mathbf{y}\|}{h} E(\mathbf{a}, \mathbf{v}).$$

Porque $\|\mathbf{v}\| \rightarrow 0$ quando $h \rightarrow 0$ é visto que $|h|/h = \pm 1$, o segundo membro de (8.11) tende para o limite $T_a(\mathbf{y})$ quando $h \rightarrow 0$. Portanto o primeiro membro tende para o mesmo limite, o que demonstra (8.8).

Façamos agora uso da linearidade de T_a para deduzir (8.9). Se $\mathbf{y} = (y_1, \dots, y_n)$ temos $\mathbf{y} = \sum_{k=1}^n y_k \mathbf{e}_k$, onde resulta

$$T_a(\mathbf{y}) = T_a\left(\sum_{k=1}^n y_k \mathbf{e}_k\right) = \sum_{k=1}^n y_k T_a(\mathbf{e}_k) = \sum_{k=1}^n y_k f'(\mathbf{a}; \mathbf{e}_k) = \sum_{k=1}^n y_k D_k f(\mathbf{a}).$$

8.12. Gradiente de um campo escalar

A fórmula do teorema 8.5, que exprime $f'(\mathbf{a}; \mathbf{y})$ como numa combinação linear das componentes de \mathbf{y} , pode escrever-se como um produto escalar

$$f'(\mathbf{a}; \mathbf{y}) = \sum_{k=1}^n D_k f(\mathbf{a}) y_k = \nabla f(\mathbf{a}) \cdot \mathbf{y},$$

onde $\nabla f(\mathbf{a})$ é o vector cujas componentes são as derivadas parciais de f em \mathbf{a} ,

$$\nabla f(\mathbf{a}) = (D_1 f(\mathbf{a}), \dots, D_n f(\mathbf{a})).$$

Este é chamado *gradiente de f* . O gradiente ∇f é um campo vectorial definido em cada ponto \mathbf{a} em que existem as derivadas parciais $D_1 f(\mathbf{a}), \dots, D_n f(\mathbf{a})$. Usa-se igualmente a notação $\text{grad } f$ em vez de ∇f . O símbolo ∇ lê-se “nabla”.

A fórmula de Taylor de primeira ordem (8.10) pode agora escrever-se na forma

$$(8.12) \quad f(\mathbf{a} + \mathbf{v}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot \mathbf{v} + \|\mathbf{v}\| E(\mathbf{a}, \mathbf{v}),$$

onde $E(\mathbf{a}, \mathbf{v}) \rightarrow 0$ quando $\|\mathbf{v}\| \rightarrow 0$. Escrita nesta forma assemelha-se à fórmula de Taylor unidimensional com o vector gradiente $\nabla f(\mathbf{a})$ desempenhando o papel da derivada $f'(\mathbf{a})$.

A partir da fórmula de Taylor podemos facilmente provar que a diferenciabilidade implica a continuidade.

TEOREMA 8.6. *Se um campo escalar f é diferenciável em \mathbf{a} , então f é contínua em \mathbf{a} .*

Demonstração. De (8.12) tem-se

$$|f(\mathbf{a} + \mathbf{v}) - f(\mathbf{a})| = |\nabla f(\mathbf{a}) \cdot \mathbf{v} + \|\mathbf{v}\| E(\mathbf{a}, \mathbf{v})|.$$

Aplicando a desigualdade triangular e a desigualdade de Cauchy-Schwarz encontramos

$$0 \leq |f(\mathbf{a} + \mathbf{v}) - f(\mathbf{a})| \leq \|\nabla f(\mathbf{a})\| \|\mathbf{v}\| + \|\mathbf{v}\| |E(\mathbf{a}, \mathbf{v})|.$$

Isto mostra que $f(\mathbf{a} + \mathbf{v}) \rightarrow f(\mathbf{a})$ quando $\|\mathbf{v}\| \rightarrow 0$, pelo que f é contínuo em \mathbf{a} .

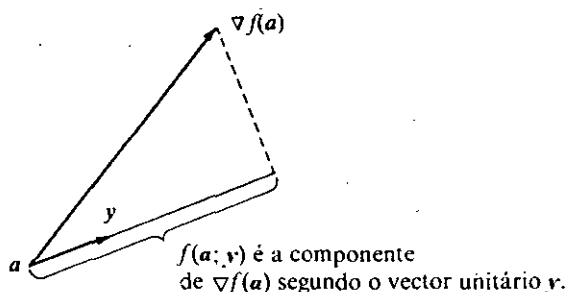


FIG. 8.5. Relação geométrica entre a derivada direccional e o vector gradiente.

Quando \mathbf{y} é um vector *unitário* a derivada direcciona $f'(\mathbf{a}; \mathbf{y})$ admite uma relação geométrica simples com o vector gradiente. Admitamos que $\nabla f(\mathbf{a}) \neq \mathbf{0}$ e seja θ ângulo de \mathbf{y} com $\nabla f(\mathbf{a})$. Então tem-se

$$f'(\mathbf{a}; \mathbf{y}) = \nabla f(\mathbf{a}) \cdot \mathbf{y} = \|\nabla f(\mathbf{a})\| \|\mathbf{y}\| \cos \theta = \|\nabla f(\mathbf{a})\| \cos \theta.$$

Isto mostra que a derivada direcciona é a componente do vector gradiente na direcção de \mathbf{y} . A figura 8.5 representa os vectores $\nabla f(\mathbf{a})$ e \mathbf{y} aplicados no ponto \mathbf{a} . A derivada é máxima quando $\cos \theta = 1$, isto é, quando \mathbf{y} tem a mesma direcção que $\nabla f(\mathbf{a})$. Por outras palavras, num dado ponto \mathbf{a} o campo escalar sofre a variação máxima na direcção do vector gradiente; além disso, este máximo é igual ao comprimento do vector gradiente. Quando $\nabla f(\mathbf{a})$ é ortogonal a \mathbf{y} , a derivada direcciona $f'(\mathbf{a}; \mathbf{y})$ é 0.

No espaço bidimensional o vector gradiente escreve-se muitas vezes

$$\nabla f(x, y) = \frac{\partial f(x, y)}{\partial x} \mathbf{i} + \frac{\partial f(x, y)}{\partial y} \mathbf{j}.$$

No espaço tridimensional a fórmula correspondente é

$$\nabla f(x, y, z) = \frac{\partial f(x, y, z)}{\partial x} \mathbf{i} + \frac{\partial f(x, y, z)}{\partial y} \mathbf{j} + \frac{\partial f(x, y, z)}{\partial z} \mathbf{k}.$$

8.13. Uma condição suficiente de diferenciabilidade

Se f é diferenciável em \mathbf{a} , então todas as derivadas parciais $D_1 f(\mathbf{a}), \dots, D_n f(\mathbf{a})$ existem. Contudo, a existência de todas estas derivadas parciais não implica necessariamente que f seja diferenciável em \mathbf{a} . Um contra exemplo é dado pela função

$$f(x, y) = \frac{xy^2}{x^2 + y^4} \quad \text{se } x \neq 0, \quad f(0, y) = 0,$$

já discutida na Secção 8.10. Para esta função, quer a derivada parcial $D_1 f(\mathbf{0})$ quer $D_2 f(\mathbf{0})$ existem, mas f não é contínua em $\mathbf{0}$, pelo que f não pode ser diferenciável em $\mathbf{0}$.

O teorema seguinte mostra que a existência de derivadas parciais *contínuas* num ponto implica a diferenciabilidade nesse ponto.

TEOREMA 8.7. UMA CONDIÇÃO SUFICIENTE DE DIFERENCIABILIDADE. *Se as derivadas parciais $D_1 f, \dots, D_n f$ existirem numa certa n -bola $B(\mathbf{a})$ e forem contínuas em \mathbf{a} , então f é diferenciável em \mathbf{a} .*

Nota: Um campo escalar satisfazendo à hipótese do teorema 8.7 diz-se *continuamente diferenciável* em \mathbf{a} .

Demonstração. $T_a(v)$ apenas pode ser $\nabla f(a) \cdot v$. Se provarmos que

$$f(a + v) - f(a) = \nabla f(a) \cdot v + \|v\| E(a, v),$$

com $E(a, v) \rightarrow 0$ quando $\|v\| \rightarrow 0$, o teorema está demonstrado.

Seja $\lambda = \|v\|$. Então $v = \lambda u$, com $\|u\| = 1$. Fazendo λ suficientemente pequeno de maneira que $a + v$ esteja na bola $B(a)$ em que existem as derivadas $D_1 f, \dots, D_n f$. Exprimindo u em função das suas componentes temos

$$u = u_1 e_1 + \dots + u_n e_n,$$

com e_1, e_2, \dots, e_n vectores unitários. Escrevamos agora a diferença $f(a + v) - f(a)$ na forma

$$(8.13) \quad f(a + v) - f(a) = f(a + \lambda u) - f(a) = \sum_{k=1}^n \{f(a + \lambda v_k) - f(a + \lambda v_{k-1})\},$$

onde v_0, v_1, \dots, v_n são vectores quaisquer de \mathbb{R}^n tais que $v_0 = O$ e $v_n = u$. Escolhamos estes vectores de maneira que verifiquem a relação $v_k = v_{k-1} + u_k e_k$.

$$v_0 = O, \quad v_1 = u_1 e_1, \quad v_2 = u_1 e_1 + u_2 e_2, \quad \dots, \quad v_n = u_1 e_1 + \dots + u_n e_n.$$

Então o termo de ordem k da soma (8.13) escreve-se

$$f(a + \lambda v_{k-1} + \lambda u_k e_k) - f(a + \lambda v_{k-1}) = f(b_k + \lambda u_k e_k) - f(b_k),$$

onde $b_k = a + \lambda v_{k-1}$. Os dois pontos b_k e $b_k + \lambda u_k e_k$ diferem unicamente nas componentes de ordem k . Deste modo podemos aplicar o teorema da média do cálculo diferencial para escrevermos

$$(8.14) \quad f(b_k + \lambda u_k e_k) - f(b_k) = \lambda u_k D_k f(c_k),$$

onde c_k pertence ao segmento definido por b_k e $b_k + \lambda u_k e_k$. Observe-se que $b_k \rightarrow a$ e por isso $c_k \rightarrow a$ quando $\lambda \rightarrow 0$.

Introduzindo (8.14) em (8.13) obtemos

$$f(a + v) - f(a) = \lambda \sum_{k=1}^n D_k f(c_k) u_k.$$

Mas $\nabla f(a) \cdot v = \lambda \nabla f(a) \cdot u = \lambda \sum_{k=1}^n D_k f(a) u_k$, pelo que

$$f(a + v) - f(a) - \nabla f(a) \cdot v = \lambda \sum_{k=1}^n \{D_k f(c_k) - D_k f(a)\} u_k = \|v\| E(a, v),$$

sendo

$$E(a, v) = \sum_{k=1}^n \{D_k f(c_k) - D_k f(a)\} u_k.$$

Visto que $c_k \rightarrow a$ quando $\|v\| \rightarrow 0$, e porque cada derivada parcial $D_k f$ é contínua em a , vê-se que $E(a, v) \rightarrow 0$ quando $\|v\| \rightarrow 0$, ficando completada a demonstração do teorema.

8.14. Exercícios

- Determinar o vector gradiente em cada ponto em que exista para os campos escalares definidos pelas equações seguintes:
 - $f(x, y) = x^2 + y^2 \sin(xy)$.
 - $f(x, y) = e^x \cos y$.
 - $f(x, y, z) = x^2 y^3 z^4$.
 - $f(x, y, z) = x^2 - y^2 + 2z^2$.
 - $f(x, y, z) = \log(x^2 + 2y^2 - 3z^2)$.
 - $f(x, y, z) = x^{2z}$.
- Calcular as derivadas direccionais dos seguintes campos escalares nos pontos e segundo as direcções indicadas.
 - $f(x, y, z) = x^2 + 2y^2 + 3z^2$ em $(1, 1, 0)$ na direcção de $\mathbf{i} - \mathbf{j} + 2\mathbf{k}$.
 - $f(x, y, z) = (x/y)^z$ em $(1, 1, 1)$ na direcção de $2\mathbf{i} + \mathbf{j} - \mathbf{k}$.
- Determinar os pontos (x, y) e as direcções segundo as quais a derivada direccionial de $f(x, y) = 3x^2 + y^2$ tem o máximo valor possível, se (x, y) está restringido aos pontos do círculo $x^2 + y^2 = 1$.
- Um campo escalar diferenciável f tem, no ponto $(1, 2)$, derivada direccionial $+2$ na direcção do ponto $(2, 2)$ e -2 na direcção do ponto $(1, 1)$. Determinar o vector gradiente em $(1, 2)$ e calcular a derivada direccionial na direcção do ponto $(4, 6)$.
- Determinar os valores das constantes a, b e c tais que a derivada direccionial de $f(x, y, z) = ax^2y + byz + cz^2x^2$ no ponto $(1, 2, -1)$ tenha o valor máximo 64 na direcção parábola ao eixo OZ .
- Dado um campo escalar diferenciável num ponto \mathbf{a} de \mathbb{R}^2 , admita-se que $f'(\mathbf{a}; \mathbf{y}) = 1$ e $f'(\mathbf{a}; \mathbf{z}) = 2$, onde $\mathbf{y} = 2\mathbf{i} + 3\mathbf{j}$ e $\mathbf{z} = \mathbf{i} + \mathbf{j}$. Traçar um gráfico mostrando o conjunto de todos os pontos (x, y) para os quais $f(\mathbf{a}; x\mathbf{i} + y\mathbf{j}) = 6$. Calcular ainda $\nabla f(\mathbf{a})$.
- Sejam f e g campos escalares diferenciáveis num conjunto aberto S . Demonstrar as seguintes propriedades do gradiente:
 - $\text{grad } f = \mathbf{0}$ se f é constante em S .
 - $\text{grad } (f + g) = \text{grad } f + \text{grad } g$.
 - $\text{grad } (cf) = c \text{ grad } f$ se c for constante.
 - $\text{grad } (fg) = \text{grad } g + g \text{ grad } f$.
 - $\text{grad } \left(\frac{f}{g} \right) = \frac{g \text{ grad } f - f \text{ grad } g}{g^2}$ em todos os pontos para os quais $g \neq 0$.
- Em \mathbb{R}^3 consideremos $\mathbf{r}(x, y, z) = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ e $r(x, y, z) = \|\mathbf{r}(x, y, z)\|$.
 - Demonstrar que $\nabla r(x, y, z)$ é um vector unitário na direcção de $\mathbf{r}(x, y, z)$.
 - Demonstrar que $\nabla(r^n) = nr^{n-2}\mathbf{r}$ se n é um inteiro positivo. [Sugestão: Utilizar o Exercício 7(d).]
 - Será válida a fórmula da alínea (b) quando n for um inteiro negativo ou zero?
 - Determinar um campo escalar f tal que $\nabla f = \mathbf{r}$.
- Supor que f é diferenciável em cada ponto de uma n -bola $B(\mathbf{a})$. Se $f'(\mathbf{x}; \mathbf{y}) = 0$ para n vectores independentes $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$ e para todo \mathbf{x} em $B(\mathbf{a})$, provar que f é constante em $B(\mathbf{a})$.
- Supor que f é diferenciável em cada ponto de uma n -bola $B(\mathbf{a})$.
 - Se $\nabla f(\mathbf{x}) = \mathbf{0}$ para todo \mathbf{x} em $B(\mathbf{a})$, provar que f é constante em $B(\mathbf{a})$.
 - Se $f(\mathbf{x}) \leq f(\mathbf{a})$ para todo \mathbf{x} em $B(\mathbf{a})$, provar que $\nabla f(\mathbf{a}) = \mathbf{0}$.
- Considerem-se as seis proposições seguintes relativas a um campo escalar $f: S \rightarrow \mathbb{R}$, onde $S \subseteq \mathbb{R}^n$ e \mathbf{a} é um ponto interior de S .
 - f é contínua em \mathbf{a} .
 - f é diferenciável em \mathbf{a} .
 - $f'(\mathbf{a}; \mathbf{y})$ existe para todo \mathbf{y} em \mathbb{R}^n .
 - Todas as derivadas parciais de primeira ordem de f existem numa vizinhança de \mathbf{a} e são contínuas em \mathbf{a} .

$$(e) \nabla f(a) = 0.$$

$$(f) f(x) = \|x - a\| \text{ para todo } x \text{ em } \mathbb{R}^n.$$

Numa tabela semelhante à representada aqui, marcar com T o quadrado correspondente se a proposição da linha (x) implica sempre a posição da coluna (y). Por exemplo, se (a) implica sempre (b) escrever T no segundo quadrado da primeira linha. A diagonal principal já foi marcada.

	a	b	c	d	e	f
a	T					
b		T				
c			T			
d				T		
e					T	
f						T

8.15. Generalização da regra de derivação de funções compostas para derivadas de campos escalares

Na teoria de derivação a uma dimensão, a regra de derivação de uma função composta permite-nos calcular a derivada de $g(t) = f[r(t)]$ pela fórmula

$$g'(t) = f'[r(t)] \cdot r'(t).$$

Vamos em seguida obter uma extensão da fórmula quando f se substitui por um campo escalar definido num conjunto do espaço n -dimensional e r é substituída por uma função vectorial de uma variável real cujos valores pertencem ao domínio de f .

Posteriormente generalizaremos a fórmula de modo a cobrir as hipóteses em que tanto f como r são campos vectoriais.

É fácil enumerar exemplos práticos nos quais se verifique a composição de um campo escalar com um campo vectorial. Por exemplo, suponhamos que $f(x)$ mede a temperatura num ponto x de um sólido no 3-espaço, e que pretendemos saber como a temperatura varia quando o ponto x se desloca ao longo da curva C definida no sólido. Se a curva é dada por uma função vectorial r definida no intervalo $[a, b]$, podemos introduzir uma nova função g mediante a fórmula

$$g(t) = f[r(t)] \quad \text{se } a \leq t \leq b.$$

Esta função composta g exprime a temperatura como uma função do parâmetro t e a sua derivada $g'(t)$ mede a variação da temperatura ao longo da curva. A seguinte generalização da regra de derivação da função composta já conhecida, permite-nos calcular a derivada $g'(t)$ sem determinar explicitamente $g(t)$.

TEOREMA 8.8. *Sejam f um campo escalar definido num conjunto aberto S de \mathbb{R}^n e r uma função vectorial que aplica um intervalo J de \mathbb{R}^1 em S . Defina-se a função composta $g = f \circ r$ por*

$$g(t) = f[r(t)] \quad \text{se } t \in J.$$

Seja t um ponto de J no qual $r'(t)$ existe e admita-se que f é diferenciável em $r(t)$. Então $g'(t)$ existe e é igual ao produto escalar

$$(8.15) \quad g'(t) = \nabla f(a) \cdot r'(t), \quad \text{com } a = r(t).$$

Demonstração. Façamos $a = r(t)$, com t um ponto de J no qual $r'(t)$ existe. Porque S é aberto existe uma n -bola $B(a)$ situada em S . Tomemos $h \neq 0$, mas suficientemente pequeno para que $r(t+h)$ esteja em $B(a)$, e seja $y = r(t+h) - r(t)$. Observe-se que $y \rightarrow 0$ quando $h \rightarrow 0$. Temos agora

$$g(t+h) - g(t) = f[r(t+h)] - f[r(t)] = f(a+y) - f(a).$$

Aplicando a fórmula de Taylor de primeira ordem para f temos

$$f(a+y) - f(a) = \nabla f(a) \cdot y + \|y\| E(a, y),$$

onde $E(a, y) \rightarrow 0$ quando $\|y\| \rightarrow 0$. Visto que $y = r(t+h) - r(t)$ temos

$$\frac{g(t+h) - g(t)}{h} = \nabla f(a) \cdot \frac{r(t+h) - r(t)}{h} + \frac{\|r(t+h) - r(t)\|}{h} E(a, y).$$

Fazendo $h \rightarrow 0$ obtemos (8.15).

EXEMPLO 1. *Derivada direcciona ao longo de uma curva.* Quando a função r descreve uma curva C , a derivada r' é o vector velocidade (tangente à curva) e a derivada g' em (8.15) é a derivada de f relativamente ao vector velocidade, supondo que $r' \neq 0$. Se $T(t)$ é um vector unitário na direcção de $r'(t)$ (T é o vector unitário tangente), o produto escalar $\Delta f[r(t)]$. $T(t)$ chama-se a *derivada direcciona de f ao longo da curva C ou na direcção de C* . Para uma curva plana podemos escrever

$$T(t) = \cos \alpha(t) i + \sin \beta(t) j,$$

onde $\alpha(t)$ e $\beta(t)$ são os ângulos definidos por $T(t)$ e pelos semi-eixos positivos OX e OY ; a derivada direcciona de f ao longo de C vem

$$\nabla f[r(t)] \cdot T(t) = D_1 f[r(t)] \cos \alpha(t) + D_2 f[r(t)] \sin \beta(t).$$

Esta fórmula escreve-se, muitas vezes, de maneira abreviada

$$\nabla f \cdot \mathbf{T} = \frac{\partial f}{\partial x} \cos \alpha + \frac{\partial f}{\partial y} \cos \beta.$$

Alguns autores escrevem df/ds para a derivada direcciona $\Delta f \cdot \mathbf{T}$. Visto que a derivada direcciona ao longo de C é definida por intermédio de \mathbf{T} , o seu valor depende da representação paramétrica escolhida para C . Uma mudança na representação poderá inverter o sentido de \mathbf{T} ; mas isso, por sua vez, inverterá o sentido da derivada direcciona.

EXEMPLO 2. Determinar a derivada direcciona do campo escalar $f(x, y) = x^2 - 3xy$ ao longo da parábola $y = x^2 - x + 2$ no ponto $(1, 2)$.

Resolução. Num ponto arbitrário (x, y) o vector gradiente é

$$\nabla f(x, y) = \frac{\partial f}{\partial x} \mathbf{i} + \frac{\partial f}{\partial y} \mathbf{j} = (2x - 3y)\mathbf{i} - 3x\mathbf{j}.$$

No ponto $(1, 2)$ tem-se $\Delta f(1, 2) = -4\mathbf{i} - 3\mathbf{j}$. A parábola define-se parametricamente pela equação vectorial $\mathbf{r}(t) = t\mathbf{i} + (t^2 - t + 2)\mathbf{j}$. Deste modo, $\mathbf{r}(1) = \mathbf{i} + 2\mathbf{j}$, $\mathbf{r}'(t) = \mathbf{i} + (2t - 1)\mathbf{j}$ e $\mathbf{r}'(1) = \mathbf{i} + \mathbf{j}$. Para esta representação de C o vector tangente unitária $\mathbf{T}(1)$ é $(\mathbf{i} + \mathbf{j})/\sqrt{2}$ e a derivada direcciona pretendida é $\Delta f(1, 2) \cdot \mathbf{T}(1) = -7/\sqrt{2}$.

EXEMPLO 3. Sejam f um campo escalar não constante, diferenciável em todo o plano, e c uma constante. Admitamos que a equação cartesiana $f(x, y) = c$ define uma curva C , admitindo uma tangente em cada um dos seus pontos. Provar que f possui as seguintes propriedades em cada ponto de C .

- (a) O vector gradiente Δf é normal a C .
- (b) A derivada direcciona de f ao longo de C é zero.
- (c) A derivada direcciona de f tem o seu valor máximo na direcção normal a C .

Resolução. Se \mathbf{T} é um vector unitário tangente a C , a derivada direcciona de f ao longo de C é o produto escalar $\Delta f \cdot \mathbf{T}$. Este produto é nulo se Δf for perpendicular a \mathbf{T} e tem o seu máximo valor se Δf for paralelo a \mathbf{T} . Deste modo, ambas as proposições (b) e (c) são consequências de (a). Para demonstrar (a), consideremos qualquer curva plana Γ com uma equação vectorial da forma $\mathbf{r}(t) = X(t)\mathbf{i} + Y(t)\mathbf{j}$ e introduzamos a função $g(t) = f(\mathbf{r}(t))$. Pela regra de derivação da função composta temos $g'(t) = \Delta f(\mathbf{r}(t)) \cdot \mathbf{r}'(t)$. Quando $\Gamma = C$, a função g tem o valor constante c pelo que $g'(t) = 0$ se $\mathbf{r}(t) \in C$. Uma vez que $g' = \Delta f \cdot \mathbf{r}'$, isto mostra que Δf é perpendicular a \mathbf{r}' em C ; logo Δf é normal a C .

8.16. Aplicações geométricas. Conjuntos de nível. Planos tangentes.

A regra de derivação da função composta pode-utilizar-se para deduzir propriedades geométricas do vector gradiente. Seja f um campo escalar definido num conjunto S de

\mathbb{R}^n e consideremos aqueles pontos x de S para os quais $f(x)$ toma um valor constante, por exemplo $f(x) = c$. Representemos este conjunto por $L(c)$, pelo que

$$L(c) = \{x \mid x \in S \text{ e } f(x) = c\}.$$

O conjunto $L(c)$ chama-se um *conjunto de nível* de f . Em \mathbb{R}^2 , $L(c)$ diz-se uma *curva de nível*; em \mathbb{R}^3 diz-se uma *superfície de nível*.

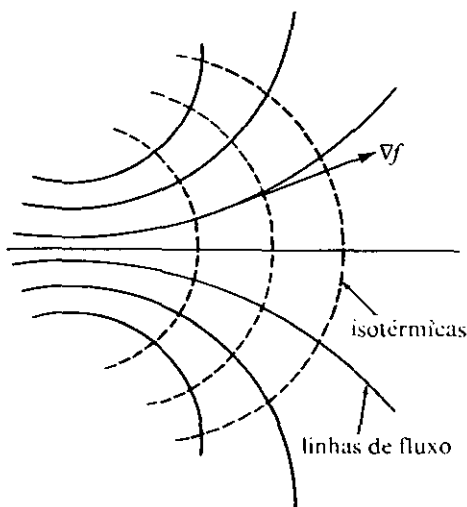


FIG. 8.6. As curvas a tracejado são isotérmicas: $f(x, y) = c$. O vector gradiente Δf indica a direcção das linhas de fluxo.

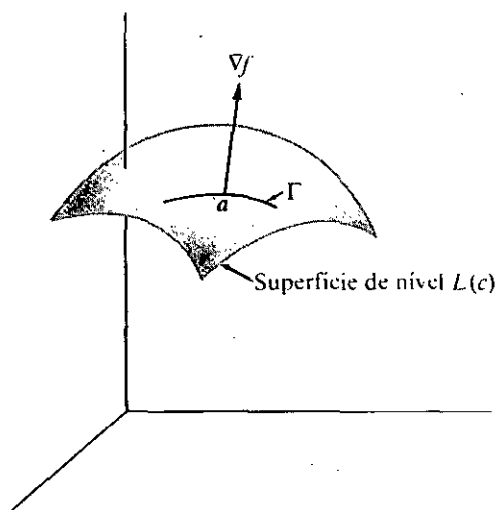


FIG. 8.7. O vector gradiente Δf é normal a cada curva Γ na superfície de nível $f(x, y, z) = c$.

Famílias de conjuntos de nível aparecem em muitas aplicações físicas. Por exemplo, se $f(x, y)$ representa a temperatura em (x, y) as curvas de nível de f (curvas de temperatura constante) chamam-se *isotérmicas*. O fluxo de calor faz-se na direcção da variação mais rápida da temperatura. Como se mostrou no Exemplo 3 da secção anterior, esta direcção é normal às isotérmicas. Logo, numa folha plana delgada o fluxo de calor faz-se ao longo da família de curvas ortogonais às isotérmicas. Aquelas são as chamadas *linhas de fluxo*; são as trajectórias ortogonais às isotérmicas. Na figura 8.6 estão representadas essas linhas.

Consideremos agora um campo escalar f , diferenciável num conjunto aberto S de \mathbb{R}^3 , e examinemos uma das suas superfícies de nível, $L(c)$. Seja a um ponto desta superfície, e consideremos uma curva Γ situada na superfície e passando por a , como se mostra na figura 8.7. Vamos demonstrar que o vector gradiente $\Delta f(a)$ é normal a esta curva em a ou, por outras palavras, vamos demonstrar que $\Delta f(a)$ é perpendicular ao vector tangente a Γ em a . Com este objectivo admitimos que é representada parametricamente

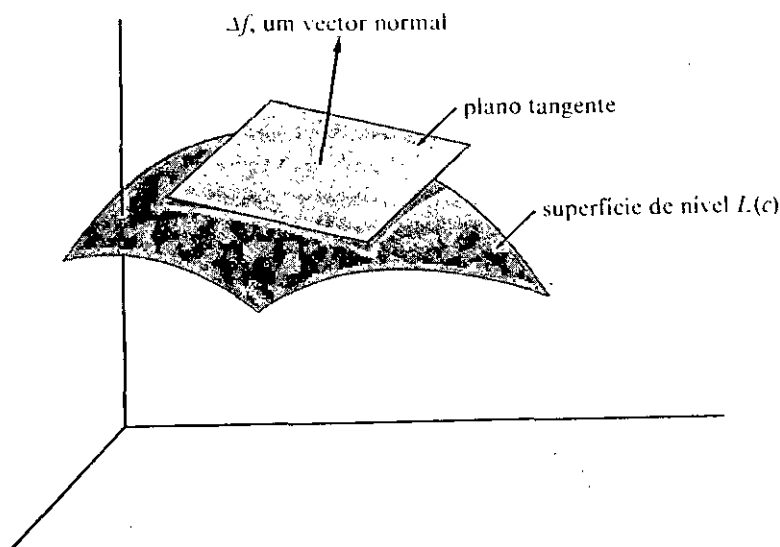


FIG. 8.8. O vector gradiente Δf é normal ao plano tangente à superfície de nível $f(x, y, z) = c$.

camente por uma função vectorial derivável r , definida em certo intervalo J de \mathbb{R}^1 . Uma vez que Γ está situada na superfície de nível $L(c)$, a função r satisfaz à equação.

$$f[r(t)] = c \text{ para todo } t \text{ em } J.$$

Se $g(t) = f[r(t)]$ para t em J , a regra de derivação da função composta estabelece que

$$g'(t) = \nabla f[r(t)] \cdot r'(t).$$

Visto que g é constante em J , temos $g'(t) = 0$ em J . Em particular, escolhendo t_1 de tal maneira que $g(t_1) = c$ encontramos que

$$\nabla f(a) \cdot r'(t_1) = 0.$$

Por outras palavras, o gradiente de f em a é perpendicular ao vector tangente $r'(t_1)$ como se afirmara.

Tomemos agora uma família de curvas na superfície de nível $L(c)$, passando todas pelo ponto a . De acordo com a discussão precedente, os vectores tangentes de todas estas curvas são perpendiculares ao vector gradiente $\Delta f(a)$. Se $\Delta f(a)$ não é um vector nulo, aqueles vectores tangentes definem um plano, e o gradiente $\Delta f(a)$ é normal a este plano. (Ver figura 8.8). Este é o *plano tangente* à superfície do nível $L(c)$ em a .

Sabemos do Volume I que um plano tangente passando por a , com vector normal N , é formado por todos os pontos x de \mathbf{R}^3 verificando $N \cdot (x - a) = 0$. Deste modo, o plano tangente à superfície de nível $L(c)$ em a será formado por todos os pontos de \mathbf{R}^3 satisfazendo a

$$\nabla f(a) \cdot (x - a) = 0.$$

Para obtermos a equação cartesiana deste plano exprimimos x , a e $\Delta f(a)$ por intermédio das suas componentes. Escrevendo $x = (x, y, z)$, $a = (x_1, y_1, z_1)$, e

$$\nabla f(a) = D_1 f(a)i + D_2 f(a)j + D_3 f(a)k,$$

obtemos a equação cartesiana

$$D_1 f(a)(x - x_1) + D_2 f(a)(y - y_1) + D_3 f(a)(z - z_1) = 0.$$

Para os campos escalares definidos em \mathbf{R}^2 é válida uma discussão análoga. No exemplo 3 da Secção anterior provámos que o vector gradiente $\Delta f(a)$, num ponto a de uma curva de nível, é perpendicular ao vector tangente à curva em a . Portanto a tangente à curva de nível $L(c)$ num ponto $a = (x_1, y_1)$ é a recta de equação cartesiana

$$D_1 f(a)(x - x_1) + D_2 f(a)(y - y_1) = 0.$$

8.17. Exercícios

1. Neste exercício deve pressupor-se a existência e continuidade de todas as derivadas que se considerem. As equações $u = f(x, y)$, $x = X(t)$, $y = Y(t)$ definem u como função de t , seja $u = F(t)$.

(a) Aplicar a regra de derivação da função composta para provar que

$$F'(t) = \frac{\partial f}{\partial x} X'(t) + \frac{\partial f}{\partial y} Y'(t),$$

onde $\partial f / \partial x$ e $\partial f / \partial y$ são calculadas em $[X(t), Y(t)]$.

(b) De uma maneira análoga, exprimir a segunda derivada $F''(t)$ em função das derivadas de f , X e Y . Recordar-se que as derivadas parciais na fórmula da alínea (a) são funções compostas dadas por

$$\frac{\partial f}{\partial x} = D_1 f[X(t), Y(t)], \quad \frac{\partial f}{\partial y} = D_2 f[X(t), Y(t)].$$

2. Tendo em conta o Exercício 1, calcular $F'(t)$ e $F''(t)$ em função de t para cada um dos seguintes casos particulares:

(a) $f(x, y) = x^2 + y^2$, $X(t) = t$, $Y(t) = t^2$.

(b) $f(x, y) = e^{xy} \cos(xy^2)$, $X(t) = \cos t$, $Y(t) = \sin t$.

(c) $f(x, y) = \log[(1 + e^{x^2})/(1 + e^{y^2})]$, $X(t) = e$, $Y(t) = e^{-t}$.

3. Em cada alínea, calcular a derivada direccional de f para os pontos e direcções que se indicam:
- (a) $f(x, y, z) = 3x - 5y + 2z$ em $(2, 2, 1)$ na direcção da normal exterior à esfera $x^2 + y^2 + z^2 = 9$.
- (b) $f(x, y, z) = x^2 - y^2$ num ponto geral da superfície $x^2 + y^2 + z^2 = 4$ na direcção da normal exterior no referido ponto.
- (c) $f(x, y, z) = x^2 + y^2 - z^2$ em $(3, 4, 5)$ ao longo da curva de intersecção de duas superfícies $2x^2 + 2y^2 - z^2 = 25$ e $x^2 + y^2 = z^2$.
4. (a) Determinar um vector $V(x, y, z)$ normal à superfície

$$z = \sqrt{x^2 + y^2} + (x^2 + y^2)^{3/2}$$

num ponto qualquer (x, y, z) da superfície, $(x, y, z) \neq (0, 0, 0)$.

- (b) Determinar o cosseno do ângulo θ entre $V(x, y, z)$ e o semi-eixo positivo OZ e determinar o limite de $\cos \theta$ quando $(x, y, z) \rightarrow (0, 0, 0)$.
5. As duas equações $e^u \cos v = x$ e $e^u \sin v = y$ definem u e v como funções de x e y , a saber $u = U(x, y)$ e $v = V(x, y)$. Determinar fórmulas explícitas para $U(x, y)$ e $V(x, y)$, válidas para $x > 0$, e provar que os vectores $\Delta U(x, y)$ e $\Delta V(x, y)$ são perpendiculares em cada ponto (x, y) .
6. Seja $f(x, y) = \sqrt{|xy|}$.
- (a) Verificar que $\partial f / \partial x$ e $\partial f / \partial y$ são ambas nulas na origem.
- (b) Admitir a superfície $z = f(x, y)$ um plano tangente na origem [*Sugestão*: Considerar a secção feita na superfície pelo plano $x = y$].
7. Se (x_0, y_0, z_0) é um ponto da superfície $z = xy$, então as duas rectas $z = y_0x$, $y = y_0$ e $z = x_0y$, $x = x_0$ intersectam-se em (x_0, y_0, z_0) e pertencem à superfície. Verificar que o plano tangente a esta superfície no ponto (x_0, y_0, z_0) contém aquelas duas rectas.
8. Determinar a equação cartesiana do plano tangente à superfície $xyz = a^3$ num ponto qualquer (x_0, y_0, z_0) . Mostrar que o volume do tetraedro limitado por este plano e os três planos coordenados é $9a^3/2$.
9. Determinar um par de equações cartesianas lineares para a recta que é tangente a ambas as superfícies $x^2 + y^2 + 2z^2 = 4$ e $z = e^{x-y}$ no ponto $(1, 1, 1)$.
10. Determinar uma constante c tal que em qualquer ponto da intersecção de duas esferas

$$(x - c)^2 + y^2 + z^2 = 3 \quad \text{e} \quad x^2 + (y - 1)^2 + z^2 = 1$$

os correspondentes planos tangentes sejam perpendiculares entre si.

11. Se r_1 e r_2 representam as distâncias de um ponto (x, y) de uma elipse aos focos, mostrar que a equação $r_1 + r_2 = \text{constante}$ (a que essas distâncias satisfazem) implica a relação

$$T \cdot \nabla(r_1 + r_2) = 0,$$

onde T é o vector tangente unitária à curva. Interpretar este resultado geometricamente e com ele demonstrar que a tangente faz ângulos iguais com as rectas unindo (x, y) aos focos.

12. Se $\Delta f(x, y, z)$ é sempre paralelo a $x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$, mostrar que f deve tomar valores iguais nos pontos $(0, 0, a)$ e $(0, 0, -a)$.

8.18. Derivadas de campos vectoriais

A teoria da derivação para campos vectoriais é uma extensão directa da mesma teoria para campos escalares. Seja $f: S \rightarrow \mathbb{R}^m$ um campo vectorial definido num subconjunto S de \mathbb{R}^n . Se a é um ponto interior de S e se y é um vector qualquer de \mathbb{R}^n , definimos a derivada $f'(a; y)$ pela fórmula

$$f'(a; y) = \lim_{h \rightarrow 0} \frac{f(a + hy) - f(a)}{h},$$

sempre que tal limite exista. A derivada $f'(a; y)$ é um vector de \mathbb{R}^m .

Designemos por f_k a componente de ordem k de f . Observemos que a derivada $f'(a; y)$ existe se e só se $f_k'(a; y)$ existe para cada $k = 1, 2, \dots, m$, em cujo caso se tem

$$f'(a; y) = (f_1'(a; y), \dots, f_m'(a; y)) = \sum_{k=1}^m f_k'(a; y) e_k,$$

em que e_k é o vector coordenado unitário de ordem k .

Dizemos que f é *diferenciável* num ponto interior a se existir uma transformação linear

$$T_a: \mathbb{R}^n \rightarrow \mathbb{R}^m$$

tal que

$$(8.16) \quad f(a + v) = f(a) + T_a(v) + \|v\| E(a, v),$$

com $E(a, v) \rightarrow 0$ quando $v \rightarrow 0$. A fórmula de Taylor de primeira ordem (8.16) é válida para todo v , com $\|v\| < r$ para algum $r > 0$. O termo $E(a, v)$ é um vector em \mathbb{R}^m . A transformação linear T_a chama-se a *diferencial total* ou simplesmente *diferencial* de f em a .

Para os campos escalares demonstrou-se que $T_a(y)$ é o produto escalar do vector gradiente $\Delta f(a)$ por y . Para campos vectoriais provaremos que $T_a(y)$ é um vector cuja componente de ordem k é o produto escalar $\Delta f_k(a) \cdot y$.

TEOREMA 8.9. *Se f é diferenciável em a com diferencial T_a então a derivada $f'(a; y)$ existe para todo a em \mathbb{R}^n e tem-se*

$$(8.17) \quad T_a(y) = f'(a; y).$$

Além disso, se $f = (f_1, \dots, f_m)$ e se $y = (y_1, \dots, y_n)$, tem-se

$$(8.18) \quad T_a(y) = \sum_{k=1}^m \nabla f_k(a) \cdot y e_k = (\nabla f_1(a) \cdot y, \dots, \nabla f_m(a) \cdot y).$$

Demonstração. Raciocinamos como no caso escalar. Se $y = \mathbf{O}$, então $f'(a; y) = \mathbf{O}$ e $T_a(\mathbf{O}) = \mathbf{O}$. Consequentemente podemos admitir que $y \neq \mathbf{O}$. Tomando $v = hy$ na fórmula de Taylor, temos

$$f(a + hy) - f(a) = T_a(hy) + \|hy\| E(a, v) = hT_a(y) + |h| \|y\| E(a, v).$$

Dividindo por h e fazendo $h \rightarrow 0$ obtemos (8.17).

Para demonstrar (8.18) basta observar que

$$f'(a; y) = \sum_{k=1}^m f'_k(a; y) e_k = \sum_{k=1}^m \nabla f_k(a) \cdot y e_k.$$

A equação (8.18) pode também escrever-se, de uma maneira mais simples, como um produto matricial

$$T_a(y) = Df(a)y,$$

sendo $Df(a)$ a matriz $m \times n$ cuja linha k é $\Delta f_k(a)$ e em que y é considerado como uma matriz coluna $n \times 1$. A matriz $Df(a)$ chama-se a *matriz jacobiana* de f em a . O seu elemento kj é a derivada parcial $D_{j_i} f_k(a)$. Assim tem-se

$$Df(a) = \begin{bmatrix} D_1 f_1(a) & D_2 f_1(a) & \cdots & D_n f_1(a) \\ D_1 f_2(a) & D_2 f_2(a) & \cdots & D_n f_2(a) \\ \vdots & \vdots & \ddots & \vdots \\ D_1 f_m(a) & D_2 f_m(a) & \cdots & D_n f_m(a) \end{bmatrix}.$$

A matriz jacobiana $Df(a)$ fica definida em cada ponto em que existam as mn derivadas parciais $D_{j_i} f_k(a)$.

A diferencial T_a representa-se também por $f'(a)$. A derivada $f'(a)$ é uma transformação linear; o jacobiano $Df(a)$ é uma representação matricial desta transformação.

A fórmula de Taylor de primeira ordem toma a forma

$$(8.19) \quad f(a + v) = f(a) + f'(a)(v) + \|v\| E(a, v),$$

onde $E(a, v) \rightarrow \mathbf{O}$ quando $v \rightarrow \mathbf{O}$. Ela é semelhante à fórmula de Taylor unidimensional. Para calcular as componentes do vector $f'(a)(v)$ podemos usar o produto matricial $Df(a)v$ ou a fórmula (8.18) do teorema 8.9.

8.19. A diferenciabilidade implica a continuidade

TEOREMA 8.10. Se um campo vectorial f é diferenciável em a , então f é contínuo em a .

Demonstração. Como no caso escalar, recorremos à fórmula de Taylor para demonstrar este teorema. Se fizermos $v \rightarrow O$ em (8.19) o termo de erro $\|v\| E(a, v) \rightarrow O$. A parte linear $f'(a)(v)$ tende igualmente para O , porque as transformações lineares são contínuas em O , e o teorema está demonstrado.

Ao chegar a este ponto é conveniente derivar uma desigualdade útil para a demonstração da regra de derivação composta que apresentamos a seguir. A desigualdade diz respeito ao campo vectorial f , diferenciável em a ; ela estabelece que

$$(8.20) \quad \|f'(a)(v)\| \leq M_f(a) \|v\|, \quad \text{onde} \quad M_f(a) = \sum_{k=1}^m \|\nabla f_k(a)\|.$$

Para sua demonstração utilizamos (8.18) juntamente com a desigualdade triangular e a desigualdade de Cauchy-Schwarz obtendo

$$\|f'(a)(v)\| = \left\| \sum_{k=1}^m \nabla f_k(a) \cdot v e_k \right\| \leq \sum_{k=1}^m |\nabla f_k(a) \cdot v| \leq \sum_{k=1}^m \|\nabla f_k(a)\| \|v\|.$$

8.20. Generalização da regra de derivação da função composta para derivadas de campos vectoriais

TEOREMA 8.11. *Sejam f e g campos vectoriais tais que a função composta $h = f \circ g$ esteja definida numa vizinhança do ponto a . Admita-se que g seja diferenciável em a , com diferencial $g'(a)$. Seja $b = g(a)$ e admita-se que f seja diferenciável em b , com diferencial $f'(b)$. Então h é diferenciável em a sendo a diferencial $h'(a)$ dada por*

$$h'(a) = f'(b) \circ g'(a),$$

composição das transformações lineares $f'(b)$ e $g'(a)$.

Demonstração. Consideremos a diferença $h(a+y) - h(a)$ para valores pequenos de $\|y\|$ e demonstremos que se obtém uma fórmula de Taylor de primeira ordem. Da definição de h resulta

$$(8.21) \quad h(a+y) - h(a) = f[g(a+y)] - f[g(a)] = f(b+v) - f(b),$$

sendo $v = g(a+y) - g(a)$. A fórmula de Taylor aplicada a $g(a+y)$ dá-nos

$$(8.22) \quad v = g'(a)(y) + \|y\| E_g(a, y), \quad \text{onde} \quad E_g(a, y) \rightarrow O \quad \text{quando} \quad y \rightarrow O$$

A fórmula de Taylor relativa a $f(b+v)$ dá-nos

$$(8.23) \quad f(b+v) - f(b) = f'(b)(v) + \|v\| E_f(b, v),$$

onde $E_f(b, v) \rightarrow O$ quando $v \rightarrow O$. Introduzindo (8.22) em (8.23) obtemos

$$(8.24) \quad f(b+v) - f(b) = f'(b)g'(a)(y) + f'(b)(\|y\| E_g(a, y)) + \|v\| E_f(b, v) \\ = f'(b)g'(a)(y) + \|y\| E(a, y),$$

onde $E(a, 0) = 0$ e

$$(8.25) \quad E(a, y) = f'(b)(E_g(a, y)) + \frac{\|v\|}{\|y\|} E_f(b, v) \quad \text{se } y \neq 0.$$

Para completar a demonstração necessitamos provar que $E(a, y) \rightarrow 0$ quando $y \rightarrow 0$.

O primeiro termo do segundo membro de (8.25) tende para 0 quando $y \rightarrow 0$ porque $E_g(a, y) \rightarrow 0$ quando $y \rightarrow 0$ e as transformações lineares são contínuas em 0.

No segundo termo do segundo membro de (8.25) o factor $E_f(b, v) \rightarrow 0$ porque $v \rightarrow 0$ quando $y \rightarrow 0$. O cociente $\|v\|/\|y\|$ permanece limitado porque, por (8.22) e (8.20), temos

$$\|v\| \leq M_g(a) \|y\| + \|y\| \|E_g(a, y)\|.$$

Portanto ambos os termos do segundo membro de (8.25) tendem para 0 quando $y \rightarrow 0$, pelo que $E(a, y) \rightarrow 0$.

Assim, de (8.24) e (8.21) obtemos a fórmula de Taylor

$$h(a+y) - h(a) = f'(b)g'(a)(y) + \|y\| E(a, y),$$

onde $E(a, y) \rightarrow 0$ quando $y \rightarrow 0$. Isto demonstra que h é diferenciável em a e que a diferencial $h'(a)$ é igual à composição $f'(b) \circ g'(a)$.

8.21. Forma matricial da regra de derivação para a composição

Seja $h = f \circ g$, onde g é diferenciável em a e f é diferenciável em $b = g(a)$. A regra de derivação para a composição indicada estabelece que

$$h'(a) = f'(b) \circ g'(a).$$

Podemos exprimir a regra por meio das matrizes jacobianas $Dh(a)$, $Df(b)$ e $Dg(a)$, as quais representam as transformações lineares $h'(a)$, $f'(b)$ e $g'(a)$ respectivamente. Porque a composição de transformações lineares corresponde à multiplicação dos correspondentes matrizes, obtemos

$$(8.26) \quad Dh(a) = Df(b) Dg(a), \quad \text{onde } b = g(a).$$

Esta é a *forma matricial da regra de derivação da função composta*. Pode também escrever-se como um conjunto de equações escalares expressando cada matriz em função dos seus elementos.

Suponhamos que $a \in \mathbb{R}^p$, $b = g(a) \in \mathbb{R}^n$ e $f(b) \in \mathbb{R}^m$. Então $h(a) \in \mathbb{R}^m$ e podemos escrever

$$\mathbf{g} = (g_1, \dots, g_n), \quad \mathbf{f} = (f_1, \dots, f_m), \quad \mathbf{h} = (h_1, \dots, h_m).$$

A matriz $D\mathbf{h}(\mathbf{a})$ é $m \times p$, $D\mathbf{f}(\mathbf{b})$ é $m \times n$ e $D\mathbf{g}(\mathbf{a})$ é $n \times p$, definidas por

$$D\mathbf{h}(\mathbf{a}) = [D_j h_i(\mathbf{a})]_{i,j=1}^{m,p}, \quad D\mathbf{f}(\mathbf{b}) = [D_k f_i(\mathbf{b})]_{i,k=1}^{m,n}, \quad D\mathbf{g}(\mathbf{a}) = [D_j g_k(\mathbf{a})]_{k,j=1}^{n,p}.$$

A equação matricial (8.26) é equivalente às mp equações escalares,

$$D_j h_i(\mathbf{a}) = \sum_{k=1}^n D_k f_i(\mathbf{b}) D_j g_k(\mathbf{a}), \quad \text{para } i = 1, 2, \dots, m \quad \text{e} \quad j = 1, 2, \dots, p.$$

Estas equações exprimem as derivadas parciais das componentes de \mathbf{h} , em função das derivadas parciais das componentes de \mathbf{f} e \mathbf{g} .

EXEMPLO 1. *Extensão da regra de derivação para campos escalares.* Admitamos que f é um campo escalar ($m = 1$). Então \mathbf{h} é também um campo escalar e existem p equações na regra da derivação, uma para cada uma das derivadas parciais de \mathbf{h} :

$$D_j h(\mathbf{a}) = \sum_{k=1}^n D_k f(\mathbf{b}) D_j g_k(\mathbf{a}), \quad \text{para } j = 1, 2, \dots, p.$$

O caso particular $p = 1$ já foi considerado na Secção 8.15. Nesta hipótese obtinha-se unicamente uma equação.

$$h'(\mathbf{a}) = \sum_{k=1}^n D_k f(\mathbf{b}) g'_k(\mathbf{a}).$$

Consideremos agora $p = 2$ e $n = 2$. Ponhamos $\mathbf{a} = (s, t)$ e $\mathbf{b} = (x, y)$. Então as componentes x e y estão relacionadas com s e t pelas equações

$$x = g_1(s, t), \quad y = g_2(s, t).$$

A regra de derivação das funções compostas dá-nos um par de equações para as derivadas parciais de h :

$$D_1 h(s, t) = D_1 f(x, y) D_1 g_1(s, t) + D_2 f(x, y) D_1 g_2(s, t),$$

$$D_2 h(s, t) = D_1 f(x, y) D_2 g_1(s, t) + D_2 f(x, y) D_2 g_2(s, t).$$

Na notação ∂ , a este par de equações pode dar-se a forma

$$(8.27) \quad \frac{\partial h}{\partial s} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial s} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial s}, \quad \frac{\partial h}{\partial t} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial t}.$$

EXEMPLO 2. Coordenadas polares. A temperatura de uma placa delgada é representada por um campo escalar f , sendo $f(x, y)$ a temperatura em (x, y) . Se definimos coordenadas polares r e θ por $x = r \cos \theta$ e $y = r \sin \theta$, a temperatura vem dada como uma função de r e θ definida por

$$\varphi(r, \theta) = f(r \cos \theta, r \sin \theta).$$

Exprimir as derivadas parciais $\partial \varphi / \partial r$ e $\partial \varphi / \partial \theta$ em função das derivadas parciais $\partial f / \partial x$ e $\partial f / \partial y$.

Resolução. Servimo-nos da regra de derivação da função composta tal como está expressa em (8.27), escrevendo (r, θ) em vez de (s, t) e φ em vez de h . As equações

$$x = r \cos \theta, \quad y = r \sin \theta$$

dão-nos

$$\frac{\partial x}{\partial r} = \cos \theta, \quad \frac{\partial y}{\partial r} = \sin \theta, \quad \frac{\partial x}{\partial \theta} = -r \sin \theta, \quad \frac{\partial y}{\partial \theta} = r \cos \theta.$$

Substituindo estas fórmulas em (8.27) obtemos

$$(8.28) \quad \frac{\partial \varphi}{\partial r} = \frac{\partial f}{\partial x} \cos \theta + \frac{\partial f}{\partial y} \sin \theta, \quad \frac{\partial \varphi}{\partial \theta} = -r \frac{\partial f}{\partial x} \sin \theta + r \frac{\partial f}{\partial y} \cos \theta.$$

Estas são as expressões pedidas para $\partial \varphi / \partial r$ e $\partial \varphi / \partial \theta$.

EXEMPLO 3. Derivadas parciais de segunda ordem. Continuar o Exemplo 2, exprimindo a derivada parcial $\partial^2 \varphi / \partial \theta^2$ em função das derivadas parciais de f .

Resolução. Começamos com a fórmula de $\partial \varphi / \partial \theta$ dada em (8.28) e derivamos relativamente a θ , considerando r como uma constante. Existem dois termos no segundo membro, cada um dos quais deve derivar-se como um produto. Obtemos assim

$$(8.29) \quad \begin{aligned} \frac{\partial^2 \varphi}{\partial \theta^2} &= -r \frac{\partial f}{\partial x} \frac{\partial (\sin \theta)}{\partial \theta} - r \sin \theta \frac{\partial}{\partial \theta} \left(\frac{\partial f}{\partial x} \right) + r \frac{\partial f}{\partial y} \frac{\partial (\cos \theta)}{\partial \theta} + r \cos \theta \frac{\partial}{\partial \theta} \left(\frac{\partial f}{\partial y} \right) \\ &= -r \cos \theta \frac{\partial f}{\partial x} - r \sin \theta \frac{\partial}{\partial \theta} \left(\frac{\partial f}{\partial x} \right) - r \sin \theta \frac{\partial f}{\partial y} + r \cos \theta \frac{\partial}{\partial \theta} \left(\frac{\partial f}{\partial y} \right). \end{aligned}$$

Para calcularmos as derivadas de $\partial f / \partial x$ e $\partial f / \partial y$ relativamente a θ devemos ter presente que, como funções de r e θ , $\partial f / \partial x$ e $\partial f / \partial y$ são dadas por funções compostas

$$\frac{\partial f}{\partial x} = D_1 f(r \cos \theta, r \sin \theta) \quad \text{e} \quad \frac{\partial f}{\partial y} = D_2 f(r \cos \theta, r \sin \theta).$$

Por conseguinte, as suas derivadas relativamente a θ têm que determinar-se de acordo com a regra para a derivação de funções compostas. Apliquemos de novo (8.27), substituindo f por $D_1 f$ para obtermos

$$\frac{\partial}{\partial \theta} \left(\frac{\partial f}{\partial x} \right) = \frac{\partial(D_1 f)}{\partial x} \frac{\partial x}{\partial \theta} + \frac{\partial(D_1 f)}{\partial y} \frac{\partial y}{\partial \theta} = \frac{\partial^2 f}{\partial x^2} (-r \sin \theta) + \frac{\partial^2 f}{\partial y \partial x} (r \cos \theta).$$

Analogamente, utilizando (8.27) com f substituído por $D_2 f$, encontramos

$$\frac{\partial}{\partial \theta} \left(\frac{\partial f}{\partial y} \right) = \frac{\partial(D_2 f)}{\partial x} \frac{\partial x}{\partial \theta} + \frac{\partial(D_2 f)}{\partial y} \frac{\partial y}{\partial \theta} = \frac{\partial^2 f}{\partial x \partial y} (-r \sin \theta) + \frac{\partial^2 f}{\partial y^2} (r \cos \theta).$$

Quando estas expressões se introduzem em (8.29) obtemos

$$\begin{aligned} \frac{\partial^2 \varphi}{\partial \theta^2} = & -r \cos \theta \frac{\partial f}{\partial x} + r^2 \sin^2 \theta \frac{\partial^2 f}{\partial x^2} - r^2 \sin \theta \cos \theta \frac{\partial^2 f}{\partial y \partial x} \\ & - r \sin \theta \frac{\partial f}{\partial y} - r^2 \sin \theta \cos \theta \frac{\partial^2 f}{\partial x \partial y} + r^2 \cos^2 \theta \frac{\partial^2 f}{\partial y^2}. \end{aligned}$$

Esta é a fórmula pretendida para $\partial^2 \varphi / \partial \theta^2$. Fórmulas análogas para as derivadas parciais $\partial^2 \varphi / \partial r^2$, $\partial^2 \varphi / (\partial r \partial \theta)$, e $\partial^2 \varphi / (\partial \theta \partial r)$ são pedidas no Exercício 5 da Secção 8.22.

8.22. Exercícios

Nestes exercícios pressupomos a diferenciabilidade de todas as funções que neles se considerem.

1. A substituição $t = g(x, y)$ transforma $F(t)$ em $f(x, y)$, onde $f(x, y) = F[g(x, y)]$.

(a) Mostrar que

$$\frac{\partial f}{\partial x} = F'[g(x, y)] \frac{\partial g}{\partial x} \quad \text{e} \quad \frac{\partial f}{\partial y} = F'[g(x, y)] \frac{\partial g}{\partial y}.$$

(b) Considerar o caso particular $F(t) = e^{\sin t}$, $g(x, y) = \cos(x^2 + y^2)$. Calcular $\partial f / \partial x$ e $\partial f / \partial y$ recorrendo às fórmulas da alínea (a). Comprovar o resultado, determinar $f(x, y)$ explicitamente em função de x e y e calcular $\partial f / \partial x$ e $\partial f / \partial y$ directamente de f .

2. A substituição $u = (x - y)/2$, $v = (x + y)/2$ muda $f(u, v)$ em $F(x, y)$. Aplicar de maneira adequada a regra de derivação da função composta para expressar as derivadas parciais $\partial F / \partial x$ e $\partial F / \partial y$ em função das derivadas parciais $\partial f / \partial u$ e $\partial f / \partial v$.

3. As equações $u = f(x, y)$, $x = X(s, t)$, $y = Y(s, t)$ definem u com uma função de s e t , digamos $u = F(s, t)$.

(a) Usar uma forma adequada da regra de derivação da função composta para exprimir as derivadas parciais $\partial F / \partial s$ e $\partial F / \partial t$ em função de $\partial f / \partial x$, $\partial f / \partial y$, $\partial X / \partial s$, $\partial X / \partial t$, $\partial Y / \partial s$, $\partial Y / \partial t$.

(b) Se $\partial^2 f / (\partial x \partial y) = \partial^2 f / (\partial y \partial x)$, mostrar que

$$\frac{\partial^2 F}{\partial s^2} = \frac{\partial f}{\partial x} \frac{\partial^2 X}{\partial s^2} + \frac{\partial^2 f}{\partial x^2} \left(\frac{\partial X}{\partial s} \right)^2 + 2 \frac{\partial X}{\partial s} \frac{\partial Y}{\partial s} \frac{\partial^2 f}{\partial x \partial y} + \frac{\partial f}{\partial y} \frac{\partial^2 Y}{\partial s^2} + \frac{\partial^2 f}{\partial y^2} \left(\frac{\partial Y}{\partial s} \right)^2.$$

- (c) Determinar fórmulas análogas para as derivadas parciais $\partial^2 F / (\partial s \partial t)$ e $\partial^2 F / \partial t^2$.
4. Resolver o Exercício 3 em cada um dos seguintes casos particulares:
- $X(s, t) = s + t$, $Y(s, t) = st$.
 - $X(s, t) = st$, $Y(s, t) = s/t$.
 - $X(s, t) = (s - t)/2$, $Y(s, t) = (s + t)/2$.
5. A introdução de coordenadas polares transforma $f(x, y)$ em $\varphi(r, \theta)$, onde $x = r \cos \theta$ e $y = r \sin \theta$. Expressar as derivadas parciais de segunda ordem $\partial^2 \varphi / \partial r^2$, $\partial^2 \varphi / (\partial r \partial \theta)$ e $\partial^2 \varphi / (\partial \theta \partial r)$ em termos das derivadas parciais de f . Podem utilizar-se as fórmulas estabelecidas no Exemplo da Secção 8.21.
6. As equações $u = f(x, y, z)$, $x = X(r, s, t)$, $y = Y(r, s, t)$ e $z = Z(r, s, t)$ definem u como uma função de r, s e t , a saber $u = F(r, s, t)$. Usar uma forma apropriada da regra de derivação da função composta para exprimir as derivadas parciais $\partial F / \partial r$, $\partial F / \partial s$ e $\partial F / \partial t$ em função das derivadas parciais de f, X, Y e Z .
7. Resolver o Exercício 6 em cada um dos seguintes casos particulares:
- $X(r, s, t) = r + s + t$, $Y(r, s, t) = r - 2s + 3t$, $Z(r, s, t) = 2r + s - t$.
 - $X(r, s, t) = r^2 + s^2 + t^2$, $Y(r, s, t) = r^2 - s^2 - t^2$, $Z(r, s, t) = r^2 - s^2 + t^2$.
8. As equações $u = f(x, y, z)$, $x = X(s, t)$, $y = Y(s, t)$, $z = Z(s, t)$ definem u como uma função de s, t , $u = F(s, t)$. Usar uma forma adequada da regra de derivação da função composta para exprimir as derivadas parciais $\partial F / \partial s$, $\partial F / \partial t$ em função das derivadas parciais de f, X, Y e Z .
9. Resolver o Exercício 8 em cada um dos seguintes casos particulares
- $X(s, t) = s^2 + t^2$, $Y(s, t) = s^2 - t^2$, $Z(s, t) = 2st$.
 - $X(s, t) = s + t$, $Y(s, t) = s - t$, $Z(s, t) = st$.
10. As equações $u = f(x, y)$, $x = X(r, s, t)$, $y = Y(r, s, t)$ definem u como uma função de r, s e t , $u = F(r, s, t)$. Usar uma forma adequada da regra de derivação de função composta para exprimir as derivadas parciais $\partial F / \partial r$, $\partial F / \partial s$ e $\partial F / \partial t$ em função das derivadas parciais de f, X e Y .
11. Resolver o Exercício 10 em cada um dos seguintes casos particulares
- $X(r, s, t) = r + s$, $Y(r, s, t) = t$.
 - $X(r, s, t) = r + s + t$, $Y(r, s, t) = r^2 + s^2 + t^2$.
 - $X(r, s, t) = r/s$, $Y(r, s, t) = s/t$.
12. Seja $h(x) = f[g(x)]$, onde $g = (g_1, \dots, g_n)$ é um campo vectorial diferenciável em a , e f é um campo escalar diferenciável em $b = g(a)$. Usar a regra de derivação da função composta para provar que o gradiente de h pode exprimir-se como uma combinação linear dos vectores gradientes das componentes de g , do modo seguinte:
- $$\nabla h(a) = \sum_{k=1}^n D_k f(b) \nabla g_k(a).$$
13. (a) Se $f(x, y, z) = xi + yj + zk$, provar que a matriz jacobiana $Df(x, y, z)$ é a matriz identidade de ordem 3.
- (b) Determinar todos os campos vectoriais diferenciáveis $f: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ para os quais a matriz jacobiana $Df(x, y, z)$ é a matriz identidade de ordem 3.
- (c) Determinar todos os campos vectoriais diferenciáveis $f: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ para os quais a matriz

jacobiana é a matriz diagonal da forma $\text{diag}(p(x), q(y), r(z))$, onde p, q e r são funções contínuas dadas.

14. Seja $f: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ e $g: \mathbb{R}^3 \rightarrow \mathbb{R}^2$ dois campos vectoriais definidos do seguinte modo:

$$f(x, y) = e^{x+2y}i + \sin(y + 2x)j,$$

$$g(u, v, w) = (u + 2v^2 + 3w^3)i + (2v - u^2)j.$$

- (a) Calcular cada uma das matrizes jacobianas $Df(x, y)$ e $Dg(u, v, w)$.
 (b) Calcular a função composta $h(u, v, w) = f[g(u, v, w)]$.
 (c) Calcular a matriz jacobiana $Dh(1, -1, 1)$.
15. Sejam $f: \mathbb{R}^3 \rightarrow \mathbb{R}^2$ e $g: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ dois campos vectoriais definidos do modo seguinte:

$$f(x, y, z) = (x^2 + y + z)i + (2x + y + z^2)j,$$

$$g(u, v, w) = uv^2w^2i + w^2\sin vj + u^2e^vk.$$

- (a) Calcular cada uma das matrizes jacobianas $Df(x, y, z)$ e $Dg(u, v, w)$.
 (b) Calcular a função composta $h(u, v, w) = f[g(u, v, w)]$.
 (c) Calcular a matriz jacobiana $Dh(u, 0, w)$.

★ 8.23. Condições suficientes para a igualdade das derivadas parciais mistas

Se f é uma função real de duas variáveis, as duas derivadas parciais mistas $D_{1,2}f$ e $D_{2,1}f$ não são necessariamente iguais. Com a notação $D_{1,2}f$ queremos significar $D_1(D_2f) = \partial^2 f / (\partial x \partial y)$ e com $D_{2,1}f$ queremos significar $D_2(D_1f) = \partial^2 f / (\partial y \partial x)$. Por exemplo, se f é definida por

$$f(x, y) = xy \frac{x^2 - y^2}{x^2 + y^2} \quad \text{para } (x, y) \neq (0, 0), \quad f(0, 0) = 0,$$

é fácil provar que $D_{2,1}f(0, 0) = -1$ e $D_{1,2}f(0, 0) = 1$. Isto pode ver-se do modo seguinte:

A definição de $D_{2,1}f(0, 0)$ estabelece que

$$(8.30) \quad D_{2,1}f(0, 0) = \lim_{k \rightarrow 0} \frac{D_1f(0, k) - D_1f(0, 0)}{k}.$$

Agora temos

$$D_1f(0, 0) = \lim_{h \rightarrow 0} \frac{f(h, 0) - f(0, 0)}{h} = 0$$

e, se $(x, y) \neq (0, 0)$, encontramos

$$D_1f(x, y) = \frac{y(x^4 + 4x^2y^2 - y^4)}{(x^2 + y^2)^2}.$$

Portanto, se $k \neq 0$ temos $D_1f(0, k) = -k^5/k^4 = -k$ e por isso

$$\frac{D_1 f(0, k) - D_1 f(0, 0)}{k} = -1.$$

Aplicando este resultado em (8.30) encontramos que $D_{2,1}f(0, 0) = -1$. Um raciocínio semelhante mostra que $D_{1,2}f(0, 0) = 1$, e por isso $D_{2,1}f(0, 0) \neq D_{1,2}f(0, 0)$.

No exemplo que acabámos de referir as duas derivadas parciais mistas $D_{1,2}f$ e $D_{2,1}f$ não são ambas contínuas na origem. Pode demonstrar-se que as duas derivadas parciais mistas são iguais num ponto (a, b) se pelo menos numa delas fôr contínua numa vizinhança desse ponto. Demonstraremos em primeiro lugar que elas são iguais se ambas forem contínuas. Mais precisamente, podemos enunciar o seguinte teorema:

TEOREMA 8.12. UMA CONDIÇÃO SUFICIENTE PARA A IGUALDADE DAS DERIVADAS PARCIAIS MISTAS: *Se f é um campo escalar tal que as derivadas parciais D_1f , D_2f , $D_{1,2}f$ e $D_{2,1}f$ existam num conjunto aberto S e se (a, b) é um ponto de S em que ambas as derivadas $D_{1,2}f$ e $D_{2,1}f$ são contínuas, tem-se*

$$(8.31) \quad D_{1,2}f(a, b) = D_{2,1}f(a, b).$$

Demonstração. Escolhamos h e k não nulos e tais que o rectângulo $R(h, k)$ com vértices (a, b) , $(a + h, b)$, $(a + h, b + k)$ e $(a, b + k)$ esteja situado em S . (Ver figura 8.9).

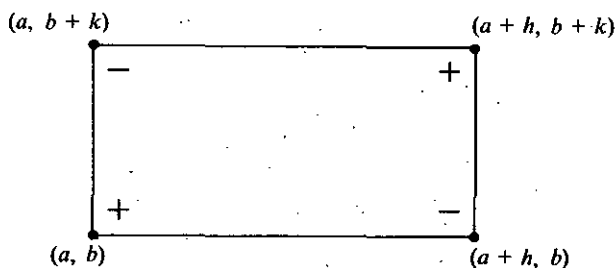


FIG. 8.9. $\Delta(h, k)$ é uma combinação de valores de f nos vértices

Consideremos a expressão

$$\Delta(h, k) = f(a + h, b + k) - f(a + h, b) - f(a, b + k) + f(a, b).$$

Esta é a combinação dos valores de f nos vértices de $R(h, k)$, tomados com os sinais indicados na figura 8.9. Devemos exprimir $\Delta(h, k)$ em função de $D_{2,1}f$ e também de $D_{1,2}f$.

Consideremos uma nova função G de uma variável definida pela equação

$$G(x) = f(x, b + k) - f(x, b)$$

para todo x compreendido entre a e $a + h$. (Geometricamente, estamos a considerar os valores de f naqueles pontos em que uma recta vertical arbitrária intersecta os lados horizontais de $R(h, k)$). Então temos

$$(8.32) \quad \Delta(h, k) = G(a + h) - G(a).$$

Aplicando o teorema da média ao segundo membro de (8.32) obtemos $G(a + h) - G(a) = hG'(x_1)$, onde x_1 está entre a e $a + h$. Uma vez que $G'(x) = D_1f(x, b + k) - D_1f(x, b)$, (8.32) escreve-se

$$(8.33) \quad \Delta(h, k) = h[D_1f(x_1, b + k) - D_1f(x_1, b)].$$

Aplicando o teorema da média ao segundo membro de (8.33) obtemos

$$(8.34) \quad \Delta(h, k) = hkD_{2,1}f(x_1, y_1),$$

em que y_1 está entre b e $b + k$. O ponto (x_1, y_1) está algures no rectângulo $R(h, k)$.

Aplicando o mesmo processo à função $H(y) = f(a + h, y) - f(a, y)$ encontramos uma segunda expressão para $\Delta(h, k)$, a saber,

$$(8.35) \quad \Delta(h, k) = hkD_{1,2}f(x_2, y_2),$$

onde (x_2, y_2) também pertence a $R(h, k)$. Igualando as duas expressões de $\Delta(h, k)$ obtemos

$$D_{1,2}f(x_1, y_1) = D_{2,1}f(x_2, y_2).$$

Fazendo agora $(h, k) \rightarrow (0, 0)$ e tendo em conta a continuidade de $D_{1,2}f$ e $D_{2,1}f$ obtemos (8.31).

O raciocínio efectuado pode modificar-se para demonstrarmos uma versão mais forte do teorema 8.12.

TEOREMA 8.13. *Se f é um campo escalar tal que as derivadas parciais D_1f , D_2f e $D_{2,1}f$ existem num conjunto aberto S contendo (a, b) e se, além disso, $D_{2,1}f$ é contínua em S , então a derivada $D_{1,2}f(a, b)$ existe e tem-se*

$$D_{1,2}f(a, b) = D_{2,1}f(a, b).$$

Demonstração. Definamos $\Delta(h, k)$ tal como na demonstração do teorema 8.12. A parte da demonstração conduzindo à equação (8.34) é ainda válida, dando-nos

$$(8.36) \quad \frac{\Delta(h, k)}{hk} = D_{2,1}f(x_1, y_1)$$

para algum (x_1, y_1) do rectângulo $R(h, k)$. O resto da demonstração não é aplicável já que é necessária a existência da derivada $D_{1,2}f(a, b)$, que é exactamente o que se pretende demonstrar.

A definição de $D_{1,2}f(a, b)$ estabelece que

$$(8.37) \quad D_{1,2}f(a, b) = \lim_{h \rightarrow 0} \frac{D_2f(a+h, b) - D_2f(a, b)}{h}.$$

Vamos provar que este limite existe e que o seu valor é $D_{2,1}f(a, b)$. Da definição de D_2f tem-se

$$D_2f(a, b) = \lim_{k \rightarrow 0} \frac{f(a, b+k) - f(a, b)}{k}$$

e

$$D_2f(a+h, b) = \lim_{k \rightarrow 0} \frac{f(a+h, b+k) - f(a+h, b)}{k}.$$

Deste modo a razão incremental em (8.37) pode escrever-se

$$\frac{D_2f(a+h, b) - D_2f(a, b)}{h} = \lim_{k \rightarrow 0} \frac{\Delta(h, k)}{hk}.$$

Tendo em conta (8.36), podemos escrever

$$(8.38) \quad \frac{D_2f(a+h, b) - D_2f(a, b)}{h} = \lim_{k \rightarrow 0} D_{2,1}f(x_1, y_1).$$

Para completarmos a demonstração temos que provar que

$$(8.39) \quad \lim_{h \rightarrow 0} \left[\lim_{k \rightarrow 0} D_{2,1}f(x_1, y_1) \right] = D_{2,1}f(a, b).$$

Quando $k \rightarrow 0$, o ponto $y_1 \rightarrow b$, mas o comportamento de x_1 como uma função de k é desconhecido. Se soubermos que x_1 se aproxima de algum limite, seja \bar{x} , quando $k \rightarrow 0$ então podemos utilizar a continuidade de $D_{2,1}f$ para provar que

$$\lim_{k \rightarrow 0} D_{2,1}f(x_1, y_1) = D_{2,1}f(\bar{x}, b).$$

Uma vez que o limite \bar{x} estará no intervalo $a \leq \bar{x} \leq a+h$, podemos então supor $h \rightarrow 0$ e deduzir (8.39). Contudo, o facto de que \bar{x} depende de k de uma maneira desconhecida torna necessário um raciocínio algo mais complicado.

Devido à equação (8.38) sabemos que existe o seguinte limite:

$$\lim_{k \rightarrow 0} D_{2,1}f(x_1, y_1).$$

Designemos este limite por $F(h)$. Para completar a demonstração teremos que

$$\lim_{h \rightarrow 0} F(h) = D_{2,1}f(a, b).$$

Para isso apelamos para a definição de continuidade de $D_{2,1}f$ em (a, b) .

Seja ϵ um número positivo dado. A continuidade de $D_{2,1}f$ em (a, b) significa que existe um disco aberto N , com centro em (a, b) e raio δ , tal que

$$(8.40) \quad |D_{2,1}f(x, y) - D_{2,1}f(a, b)| < \frac{\epsilon}{2} \quad \text{sempre que } (x, y) \in N.$$

Se escolhermos h e k de modo que $|h| < \delta/2$ e $|k| < \delta/2$, todo o rectângulo representado na figura 8.9 estará contido na vizinhança N e, em particular, o ponto (x_1, y_1) estará em N . Portanto (8.40) é verdadeiro quando $(x, y) = (x_1, y_1)$ e podemos escrever

$$(8.41) \quad 0 \leq |D_{2,1}f(x_1, y_1) - D_{2,1}f(a, b)| < \frac{\epsilon}{2}.$$

Fixemos h e façamos $k \rightarrow 0$. O termo $D_{2,1}f(x_1, y_1)$ tende para $F(h)$ e os outros termos em (8.41) são independentes de k . Temos portanto

$$0 \leq |F(h) - D_{2,1}f(a, b)| \leq \frac{\epsilon}{2} < \epsilon,$$

com tanto que $0 < |h| < \delta/2$. Mas este é precisamente o significado da proposição

$$\lim_{h \rightarrow 0} F(h) = D_{2,1}f(a, b)$$

e, como já foi dito, isto completa a demonstração.

Nota: Deve ter-se presente que o teorema é também válido se permutarmos os papéis desempenhados pelas duas derivadas $D_{1,2}f$ e $D_{2,1}f$.

8.24. Exercícios variados

- Determinar um campo escalar f que satisfaça às seguintes condições:
 - As derivadas parciais $D_1f(0, 0)$ e $D_2f(0, 0)$ existem e são nulas.
 - A derivada direcciona na origem e na direcção do vector $\mathbf{i} + \mathbf{j}$ existe e tem valor 3.
 Explicar por que razão tal função f não pode ser diferenciável em $(0, 0)$.
- Seja f a função definida por

$$f(x, y) = y \frac{x^2 - y^2}{x^2 + y^2} \quad \text{se } (x, y) \neq (0, 0), \quad f(0, 0) = 0.$$

Calcular, quando existam as seguintes derivadas parciais: $D_1f(0, 0)$, $D_2f(0, 0)$, $D_{2,1}f(0, 0)$, $D_{1,2}f(0, 0)$.

3. Seja $f(x, y) = \frac{xy^3}{x^3 + y^6}$ se $(x, y) \neq (0, 0)$, e $f(0, 0) = 0$.

(a) Provar que a derivada $f'(0; a)$ existe para cada vector a e calcular o seu valor em função das componentes de a .

(b) Determinar se f é ou não contínua na origem.

4. Define-se $f(x, y) = \int \sqrt{xy} e^{-t^2} dt$ para $x > 0, y > 0$. Calcular $\partial f / \partial x$ em função de x e y .

5. Supor que as equações $u = f(x, y)$, $x = X(t)$, $y = Y(t)$ definem u como função de t , $u = F(t)$. Calcular a terceira derivada $F'''(t)$ em função das derivadas de f , X , e Y .

6. A mudança de variáveis $x = u + v$, $y = uv^2$ transforma $f(x, y)$ em $g(u, v)$. Calcular o valor de $\partial^2 g / (\partial v \partial u)$ no ponto em que $u = 1, v = 1$, sabendo que

$$\frac{\partial f}{\partial y} = \frac{\partial^2 f}{\partial x^2} = \frac{\partial^2 f}{\partial y^2} = \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = 1$$

nesse ponto.

7. A mudança de variáveis $x = uv$, $y = \frac{1}{2}(u^2 - v^2)$ transforma $f(x, y)$ em $g(u, v)$.

(a) Calcular $\partial g / \partial u$, $\partial g / \partial v$ e $\partial^2 g / (\partial u \partial v)$ em função das derivadas parciais de f . (Pode supor-se a igualdade das derivadas parciais mistas).

(b) Se $\|\nabla f(x, y)\|^2 = 2$ para todo o par x e y , determinar as constantes a e b tais que

$$a \left(\frac{\partial g}{\partial u} \right)^2 - b \left(\frac{\partial g}{\partial v} \right)^2 = u^2 + v^2.$$

8. Duas funções F e G de uma variável e a função z de duas variáveis estão relacionadas pela equação

$$[F(x) + G(y)]^2 e^{z(x, y)} = 2F'(x)G'(y)$$

sempre que $F(x) + G(y) \neq 0$. Mostrar que a derivada parcial mista $D_{2,1}z(x, y)$ nunca se anula (Pode admitir-se a existência e continuidade de todas as derivadas calculadas).

9. Um campo escalar f é limitado e contínuo num rectângulo $R = [a, b] \times [c, d]$. Em R define-se um novo campo escalar do modo seguinte

$$g(u, v) = \int_c^v \left[\int_a^u f(x, y) dx \right] dy.$$

(a) Pode demonstrar-se que para cada u fixo em $[a, b]$ a função A definida em $[c, d]$ pela equação $A(y) = \int_a^u f(x, y) dx$ é contínua em $[c, d]$. Utilizar este resultado para demonstrar que $\partial g / \partial v$ existe e é contínuo no rectângulo aberto $S = (a, b) \times (c, d)$ (o interior de R).

(b) Supor que

$$\int_c^v \left[\int_a^u f(x, y) dx \right] dy = \int_a^u \left[\int_c^v f(x, y) dy \right] dx$$

para todo (u, v) em R . Provar que g é diferenciável em S e que as derivadas parciais mistas $D_{1,2}g(u, v)$ e $D_{2,1}g(u, v)$ existem e são iguais a $f(u, v)$ em cada ponto de S .

10. Considerar o Exercício 9. Supor u e v expressas parametricamente do modo seguinte: $u = A(t)$, $v = B(t)$; e seja $\varphi(t) = g[A(t), B(t)]$.

(a) Determinar $\varphi'(t)$ em função de f , A' e B' .

(b) Calcular $\varphi'(t)$ em função de t quando $f(x, y) = e^{x+y}$ e $A(t) = B(t) = t^2$. (Supor que R está situado no primeiro quadrante).

11. Se $f(x, y, z) = (\mathbf{r} \times \mathbf{A}) \cdot (\mathbf{r} \times \mathbf{B})$, com $\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ e \mathbf{A} e \mathbf{B} vectores constantes, mostrar que $\nabla f(x, y, z) = \mathbf{B} \times (\mathbf{r} \times \mathbf{A}) + \mathbf{A} \times (\mathbf{r} \times \mathbf{B})$.

12. Seja $\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ e $r = \|\mathbf{r}\|$. Se \mathbf{A} e \mathbf{B} são vectores constantes, mostrar que:

$$(a) \quad \mathbf{A} \cdot \nabla \left(\frac{1}{r} \right) = - \frac{\mathbf{A} \cdot \mathbf{r}}{r^3}.$$

$$(b) \quad \mathbf{B} \cdot \nabla \left(\mathbf{A} \cdot \nabla \left(\frac{1}{r} \right) \right) = \frac{3\mathbf{A} \cdot \mathbf{r} \mathbf{B} \cdot \mathbf{r}}{r^5} - \frac{\mathbf{A} \cdot \mathbf{B}}{r^3}.$$

13. Determinar o conjunto de todos os pontos (a, b, c) no 3-espaco, para os quais as duas esferas $(x-a)^2 + (y-b)^2 + (z-c)^2 = 1$ e $x^2 + y^2 + z^2 = 1$ se intersectam ortogonalmente. (Os respectivos planos tangentes deverão ser perpendiculares em cada ponto da intersecção).

14. Um cilindro de equação $y = f(x)$ é tangente à superfície $z^2 + 2xz + y = 0$ em todos os pontos comuns às duas superfícies. Determinar $f(x)$.

APLICAÇÕES DO CÁLCULO DIFERENCIAL

9.1. Equações de derivadas parciais

Os teoremas de cálculo diferencial desenvolvidos no Capítulo 8 admitem uma larga variedade de aplicações. Este capítulo ilustra o seu uso em alguns exemplos relacionados com equações de derivadas parciais, funções implícitas e problemas de extremos. Começamos com algumas notas elementares relativas às equações de derivadas parciais.

Uma equação relacionando um campo escalar f e as suas derivadas parciais chama-se uma *equação de derivadas parciais*. Dois exemplos simples, nos quais f é uma função de duas variáveis, são a equação de primeira ordem

$$(9.1) \quad \frac{\partial f(x, y)}{\partial x} = 0,$$

e a equação de segunda ordem

$$(9.2) \quad \frac{\partial^2 f(x, y)}{\partial x^2} + \frac{\partial^2 f(x, y)}{\partial y^2} = 0.$$

Cada uma delas é uma equação de derivadas parciais *linear* e homogênea, isto é, cada uma tem a forma $L(f) = 0$, onde L é um operador diferencial linear contendo uma ou mais derivadas parciais. A equação 9.2 é a *equação de Laplace* bidimensional.

Parte da teoria das equações diferenciais ordinárias lineares pode generalizar-se às equações de derivadas parciais. Por exemplo, é fácil verificar que para cada uma das equações (9.1) e (9.2) o conjunto das soluções é um espaço linear. Todavia, existe uma diferença importante entre equações lineares diferenciais ordinárias e de derivadas parciais a qual deve ser mencionada desde o princípio. Evidenciaremos esta diferença

comparando a equação de derivadas parciais (9.1) com a equação diferencial ordinária

$$(9.3) \quad f'(x) = 0.$$

A função mais geral satisfazendo a (9.3) é $f(x) = C$, com C uma constante arbitrária. Por outras palavras, o espaço-solução de (9.3) é unidimensional. Mas a função mais geral satisfazendo a (9.1) é

$$f(x, y) = g(y),$$

onde g é qualquer função de y . Visto que g é arbitrária, podemos facilmente obter um conjunto infinito de soluções independentes. Por exemplo, podemos tomar $g(y) = e^{cy}$ e fazer com que c varie no campo real. Assim, o espaço-solução de (9.1) é de dimensão infinita.

Em certos aspectos este exemplo é típico do que acontece em geral. Em dada altura, durante o processo de resolução de uma equação de derivadas parciais de primeira ordem, necessita-se de uma integração para fazer desaparecer cada derivada parcial. Nesta fase introduz-se uma função arbitrária na solução. Isto ocorre num espaço solução de dimensão infinita.

Em muitos problemas fazendo intervir equações de derivadas parciais é necessário escolher, do conjunto de soluções, uma solução particular que satisfaça a uma ou mais condições auxiliares. Como é de supor, a natureza destas condições tem um efeito fundamental na existência ou na unicidade das soluções. Um estudo sistemático de tais problemas não será efectuado neste livro. Em seu lugar, trataremos alguns casos particulares para ilustrar as ideias introduzidas no Capítulo 8.

9.2. Uma equação de derivadas parciais de primeira ordem com coeficientes constantes

Consideremos a equação de derivadas parciais de primeira ordem

$$(9.4) \quad 3 \frac{\partial f(x, y)}{\partial x} + 2 \frac{\partial f(x, y)}{\partial y} = 0.$$

Todas as soluções desta equação podem ser determinadas mediante considerações de natureza geométrica. Expressimos o primeiro membro como num produto escalar, e escrevemos a equação na forma

$$(3i + 2j) \cdot \nabla f(x, y) = 0.$$

Esta relação diz-nos que o vector gradiente $\nabla f(x, y)$ é ortogonal ao vector $3i + 2j$ em cada ponto (x, y) . Mas sabemos também que $\nabla f(x, y)$ é ortogonal às curvas de nível de f . Logo essas curvas de nível devem ser rectas paralelas a $3i + 2j$. Por outras palavras, as curvas de nível de f são as rectas

$$2x - 3y = c.$$

Deste modo $f(x, y)$ é constante quando $2x - 3y$ o fôr. Isto sugere que

$$(9.5) \quad f(x, y) = g(2x - 3y)$$

para alguma função g .

Verifiquemos agora que, para toda a função diferencial g , o campo escalar f definido por (9.5) deve, na realidade, satisfazer a (9.4). Utilizando a regra da derivação de funções compostas, para calcular as derivadas parciais de f , encontramos

$$\begin{aligned} \frac{\partial f}{\partial x} &= 2g'(2x - 3y), & \frac{\partial f}{\partial y} &= -3g'(2x - 3y), \\ 3 \frac{\partial f}{\partial x} + 2 \frac{\partial f}{\partial y} &= 6g'(2x - 3y) - 6g'(2x - 3y) = 0. \end{aligned}$$

Consequentemente, f satisfaz a (9.4).

Inversamente, podemos provar que toda a função diferenciável f que satisfaz a (9.4) terá necessariamente a forma (9.5) para algum g . Para o conseguirmos, introduzimos uma mudança de variáveis lineares,

$$(9.6) \quad x = Au + Bv, \quad y = Cu + Dv.$$

Por seu intermédio a função $f(x, y)$ transforma-se numa função de u e v , por exemplo

$$h(u, v) = f(Au + Bv, Cu + Dv).$$

Escolheremos as constantes A, B, C, D de maneira que h satisfaça à equação mais simples

$$(9.7) \quad \frac{\partial h(u, v)}{\partial u} = 0.$$

Resolveremos depois esta equação e mostraremos que f tem a forma desejada.

Aplicando a regra de derivação da função composta encontramos

$$\frac{\partial h}{\partial u} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial u} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial u} = \frac{\partial f}{\partial x} A + \frac{\partial f}{\partial y} C.$$

Visto que f satisfaz a (9.4), temos $\partial f / \partial y = -(3/2) (\partial f / \partial x)$, pelo que a equação para $\partial f / \partial u$ toma a forma

$$\frac{\partial h}{\partial u} = \frac{\partial f}{\partial x} \left(A - \frac{3}{2} C \right).$$

Por conseguinte, h verificará (9.7) se escolhermos $A = \frac{1}{2}C$. Fazendo $A = 3$ e $C = 2$ encontramos

$$(9.8) \quad x = 3u + Bv, \quad y = 2u + Dv.$$

Para esta escolha de A e C , a função h satisfaz a (9.7), pelo que $h(u, v)$ é unicamente uma função de v , por exemplo

$$h(u, v) = g(v)$$

para alguma função g . Para exprimirmos v em função de x e y eliminamos u de (9.8) e obtemos $2x - 3y = (2B - 3D)v$. Escolhemos agora B e D , de maneira que $2B - 3D = 1$, por exemplo $B = 2$, $D = 1$. Para esta escolha a transformação (9.6) é não singular; temos $v = 2x - 3y$ e portanto

$$f(x, y) = h(u, v) = g(v) = g(2x - 3y).$$

o que prova que toda a solução diferenciável f de (9.4) tem a forma (9.5).

Exactamente o mesmo tipo de raciocínio serve para a demonstração do seguinte teorema relativo a equações de derivadas parciais de primeira ordem com coeficientes constantes.

TEOREMA 9.1. *Se g é uma função diferenciável em \mathbb{R}^1 , e f é um campo escalar definido em \mathbb{R}^2 pela equação*

$$(9.9) \quad f(x, y) = g(bx - ay),$$

onde a e b são constantes não ambas nulas, então f satisfaz à equação de derivadas de primeira ordem

$$(9.10) \quad a \frac{\partial f(x, y)}{\partial x} + b \frac{\partial f(x, y)}{\partial y} = 0$$

em todo \mathbb{R}^2 . Reciprocamente, toda a solução diferenciável de (9.10) tem necessariamente a forma (9.9) para certa função g .

9.3. Exercícios

No conjunto de Exercícios que se apresenta a seguir, deve admitir-se diferenciabilidade de todas as funções que se considerem.

1. Determinar a solução da equação de derivadas parciais

$$4 \frac{\partial f(x, y)}{\partial x} + 3 \frac{\partial f(x, y)}{\partial y} = 0$$

que satisfaça à condição $f(x, 0) = \sin x$, para todo x .

2. Determinar a solução da equação de derivadas parciais

$$5 \frac{\partial f(x, y)}{\partial x} - 2 \frac{\partial f(x, y)}{\partial y} = 0$$

que satisfaça à condição $f(0, 0) = 0$ e $D_1 f(x, 0) = e^x$, para todo x .

3. (a) Se $u(x, y) = f(x, y)$, provar que u satisfaz à equação de derivadas parciais

$$x \frac{\partial u}{\partial x} - y \frac{\partial u}{\partial y} = 0.$$

Determinar uma solução tal que $u(x, x) = x^4 e^{x^2}$, para todo x .

- (b) Se $v(x, y) = f(x/y)$ para $y \neq 0$, provar que v satisfaz à equação de derivadas parciais

$$x \frac{\partial v}{\partial x} + y \frac{\partial v}{\partial y} = 0.$$

Determinar uma solução tal que $v(1, 1) = 2$ e $D_1 v(x, 1/x) = 1/x$ para todo $x \neq 0$.

4. Se $g(u, v)$ satisfaz à equação de derivadas parciais

$$\frac{\partial^2 g(u, v)}{\partial u \partial v} = 0,$$

provar que $g(u, v) = \varphi_1(u) + \varphi_2(v)$, com $\varphi_1(u)$ e $\varphi_2(v)$ unicamente funções de u e v , respectivamente.

5. Supor que f satisfaz à equação de derivadas parciais

$$\frac{\partial^2 f}{\partial x^2} - 2 \frac{\partial^2 f}{\partial x \partial y} - 3 \frac{\partial^2 f}{\partial y^2} = 0.$$

Introduzir a mudança linear de variáveis, $x = Au + Bv$, $y = Cu + Dv$, onde A, B, C e D são constantes, e seja $g(u, v) = f(Au + Bu, Cu + Dv)$. Calcular valores inteiros, não nulos, de A, B, C, D , tais que g satisfaça $\partial^2 g / (\partial u \partial v) = 0$. Resolver esta equação para g e por seu intermédio determinar f . (Supor a igualdade das derivadas parciais mistas).

6. Uma função u define-se por uma equação da forma

$$u(x, y) = xy f\left(\frac{x+y}{xy}\right).$$

Mostrar que u satisfaz à equação de derivadas parciais da forma

$$x^2 \frac{\partial u}{\partial x} - y^2 \frac{\partial u}{\partial y} = G(x, y)u,$$

e determinar $G(x, y)$.

7. A substituição $x = e^s$, $y = e^t$ transforma $f(x, y)$ em $g(s, t)$, sendo $g(s, t) = f(e^s, e^t)$. Admitindo que f satisfaz à equação de derivadas parciais

$$x^2 \frac{\partial^2 f}{\partial x^2} + y^2 \frac{\partial^2 f}{\partial y^2} + x \frac{\partial f}{\partial x} + y \frac{\partial f}{\partial y} = 0,$$

mostrar que g satisfaz à equação de derivadas parciais

$$\frac{\partial^2 g}{\partial s^2} + \frac{\partial^2 g}{\partial t^2} = 0.$$

8. Seja f um campo escalar diferenciável num conjunto aberto S de \mathbf{R}^n . Dizemos que f é *homogéneo de grau p* em S se

$$f(tx) = t^p f(x)$$

para todo $t > 0$ e cada x em S para o qual $tx \in S$. Para um campo escalar homogéneo de grau p provar que se tem

$$x \cdot \nabla f(x) = p f(x) \quad \text{para todo } x \text{ em } S.$$

Este é o chamado *teorema de Euler sobre funções homogéneas*. Se $x = (x_1, \dots, x_n)$ pode enunciar-se do seguinte modo

$$x_1 \frac{\partial f}{\partial x_1} + \dots + x_n \frac{\partial f}{\partial x_n} = p f(x_1, \dots, x_n).$$

[Sugestão: para x fixo, definir $g(t) = f(tx)$ e calcular $g'(t)$].

9. Demonstrar o recíproco do teorema de Euler, isto é, se f satisfaz a $x \cdot \nabla f(x) = p f(x)$ para todo x pertencente a um aberto S , então f é homogénea de grau p em S . [Sugestão: Para x fixo, definir $g(t) = f(tx) - t^p f(x)$ e calcular $g'(t)$].
10. Demonstrar a seguinte generalização do teorema de Euler para funções homogéneas de grau p no caso bidimensional (Supor a igualdade das derivadas parciais mistas).

$$x^2 \frac{\partial^2 f}{\partial x^2} + 2xy \frac{\partial^2 f}{\partial x \partial y} + y^2 \frac{\partial^2 f}{\partial y^2} = p(p-1)f.$$

9.4. A equação unidimensional das ondas

Imaginemos uma corda de comprimento infinito esticada ao longo do eixo OX e podendo vibrar no plano XOY . Representamos por $y = f(x, t)$ o deslocamento da corda, paralelo a a , no ponto x e no instante t . Suponhamos que, no instante $t = 0$, a corda é deslocada de modo a tomar a forma de uma curva dada $y = F(x)$. Na figura 9.1(a) apresenta-se um exemplo. As figuras 9.1(b) e 9.1(c) mostram possíveis curvas de deslocamentos para posteriores valores de t . Consideremos o deslocamento $f(x, t)$ como uma função desconhecida de x e t , a determinar. Um modelo matemático para este problema (sugerido por considerações de carácter físico que não serão analisados aqui) é a equação de derivadas parciais

$$\frac{\partial^2 f}{\partial t^2} = c^2 \frac{\partial^2 f}{\partial x^2},$$

em que c é uma constante positiva dependendo das características físicas da corda. Esta equação designa-se por *equação das ondas unidimensional*. Resolveremos esta equação tendo em conta certas condições auxiliares.

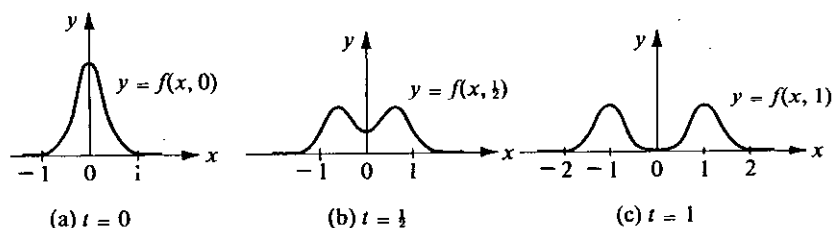


FIG. 9.1. A curva de deslocamento $y = f(x, y)$ para diferentes valores de t .

Visto que o deslocamento inicial é a curva dada $y = F(x)$, teremos que determinar uma solução satisfazendo à condição.

$$f(x, 0) = F(x).$$

Admitimos igualmente que $\partial y / \partial t$, a velocidade do deslocamento paralela OY , é definida para o instante $t = 0$, a saber

$$D_2 f(x, 0) = G(x),$$

sendo G uma função dada. Parece razoável pensar que esta informação seja suficiente para determinar o movimento subsequente da corda. Mostraremos, na verdade, que assim é, determinando a função f por intermédio de F e G . A solução exprime-se numa forma dada por Jean d'Alembert (1717-1783), um matemático e filósofo francês.

TEOREMA 9.2. SOLUÇÃO DE D'ALEMBERT DA EQUAÇÃO DAS ONDAS. *Sejam F e G funções dadas tais que G seja derivável e F duas vezes derivável em \mathbb{R}^1 . A função f definida por*

$$(9.11) \quad f(x, t) = \frac{F(x + ct) + F(x - ct)}{2} + \frac{1}{2c} \int_{x-ct}^{x+ct} G(s) ds$$

satisfaz à equação das ondas a uma dimensão

$$(9.12) \quad \frac{\partial^2 f}{\partial t^2} = c^2 \frac{\partial^2 f}{\partial x^2}$$

e às condições iniciais

$$(9.13) \quad f(x, 0) = F(x), \quad D_x f(x, 0) = G(x).$$

Inversamente, qualquer função f , com derivadas parciais mistas iguais, que satisfaça a (9.12) e (9.13) necessariamente tem a forma (9.11).

Demonstração. É um exercício imediato verificar que a função f definida por (9.11) satisfaz à equação das ondas e às condições iniciais, deixando-se tal verificação ao cuidado do leitor. Demonstraremos a inversa.

Uma maneira de proceder consiste em supôr que f é uma solução da equação das ondas, introduzir uma mudança linear de variáveis,

$$x = Au + Bv, \quad t = Cu + Dv,$$

que transforma $f(x, t)$ numa função de u e v ,

$$g(u, v) = f(Au + Bv, Cu + Dv),$$

e escolher as constantes A, B, C, D de maneira tal que g satisfaça à equação mais simples

$$\frac{\partial^2 g}{\partial u \partial v} = 0.$$

Resolvendo esta equação encontramos $g(u, v) = \varphi_1(u) + \varphi_2(v)$, onde $\varphi_1(u)$ é apenas função de u e $\varphi_2(v)$ é unicamente função de v . As constantes A, B, C, D escolhem-se de modo que $u = x + ct$, $v = x - ct$, donde se obtêm

$$(9.14) \quad f(x, t) = \varphi_1(x + ct) + \varphi_2(x - ct).$$

Em seguida usamos as condições iniciais (9.13) para determinarmos as funções φ_1 e φ_2 por intermédio das funções dadas F e G .

Obtemos (9.14) por outro método que utiliza o teorema 9.1 e evita a mudança de variáveis. Comencemos, em primeiro lugar, por escrever a equação das ondas na forma

$$(9.15) \quad L_1(L_2 f) = 0,$$

onde L_1 e L_2 são operadores diferenciais lineares de primeira ordem, definidos por

$$L_1 = \frac{\partial}{\partial t} - c \frac{\partial}{\partial x}, \quad L_2 = \frac{\partial}{\partial t} + c \frac{\partial}{\partial x}.$$

Seja f uma solução de (9.15) e

$$u(x, t) = L_2 f(x, t).$$

A equação (9.15) estabelece que u satisfaz à equação de primeira ordem $L_1(u) = 0$. Logo, pelo teorema 9.1 temos

$$u(x, t) = \varphi(x + ct)$$

para uma certa função φ . Seja Φ qualquer primitiva de φ , $\Phi(y) = \int_0^y \varphi(s) ds$, e ponhamos

$$v(x, t) = \frac{1}{2c} \Phi(x + ct).$$

Vamos demonstrar que $L_2(v) = L_2(f)$. Temos

$$\frac{\partial v}{\partial x} = \frac{1}{2c} \Phi'(x + ct) \quad \text{e} \quad \frac{\partial v}{\partial t} = \frac{1}{2} \Phi'(x + ct),$$

pelo que

$$L_2 v = \frac{\partial v}{\partial t} + c \frac{\partial v}{\partial x} = \Phi'(x + ct) = \varphi(x + ct) = u(x, t) = L_2 f.$$

Quer isto dizer que a diferença $f - v$ satisfaz à equação de primeira ordem

$$L_2(f - v) = 0.$$

Pelo teorema 9.1 deve ter-se $f(x, y) - v(x, t) = \psi(x - ct)$ para certa função ψ . Por conseguinte

$$f(x, t) = v(x, t) + \psi(x - ct) = \frac{1}{2c} \Phi(x + ct) + \psi(x - ct).$$

estando assim demonstrada (9.14), com $\varphi_1 = \frac{1}{2c} \Phi$ e $\varphi_2 = \psi$.

Façamos agora uso das condições iniciais (9.13) para determinarmos as funções φ_1 e φ_2 por intermédio das funções dadas F e G . A relação $f(x, 0) = F(x)$ implica

$$(9.16) \quad \varphi_1(x) + \varphi_2(x) = F(x).$$

A outra condição inicial, $D_2 f(x, 0) = G(x)$, implica

$$(9.17) \quad c\varphi_1'(x) - c\varphi_2'(x) = G(x).$$

Derivando (9.16) obtemos

$$(9.18) \quad \varphi_1'(x) + \varphi_2'(x) = F'(x).$$

Resolvendo (9.17) e (9.18) em relação a $\varphi_1'(x)$ e $\varphi_2'(x)$ encontramos

$$\varphi_1'(x) = \frac{1}{2} F'(x) + \frac{1}{2c} G(x), \quad \varphi_2'(x) = \frac{1}{2} F'(x) - \frac{1}{2c} G(x).$$

Integrando estas igualdades temos

$$\varphi_1(x) - \varphi_1(0) = \frac{F(x) - F(0)}{2} + \frac{1}{2c} \int_0^x G(s) ds,$$

$$\varphi_2(x) - \varphi_2(0) = \frac{F(x) - F(0)}{2} - \frac{1}{2c} \int_0^x G(s) ds.$$

Na primeira equação substituímos x por $x + ct$; na segunda substituímos x por $x - ct$. Somando as equações resultantes, membro a membro, e considerando que $\varphi_1(0) + \varphi_2(0) = F(0)$, obtemos

$$f(x, t) = \varphi_1(x + ct) + \varphi_2(x - ct) = \frac{F(x + ct) + F(x - ct)}{2} + \frac{1}{2c} \int_{x-ct}^{x+ct} G(s) ds.$$

o que completa a demonstração.

EXEMPLO. Suponhamos o deslocamento inicial dado por

$$F(x) = \begin{cases} 1 + \cos \pi x & \text{para } -1 \leq x \leq 1, \\ 0 & \text{para } |x| \geq 1. \end{cases}$$

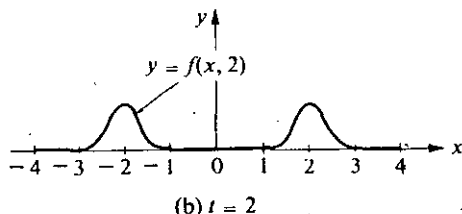
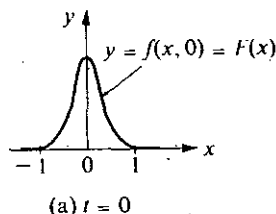


FIG. 9.2. Solução da equação das ondas, representada para $t = 0$ e $t = 2$.

O gráfico de F está traçado nas figuras 9.1(a) e 9.2(a). Admitamos que a velocidade inicial $G(x) = 0$, para todo x . Então a solução das ondas é dada pela fórmula

$$f(x, t) = \frac{F(x + ct) + F(x - ct)}{2}.$$

As figuras 9.1 e 9.2 mostram a curva $y = f(x, t)$ para diferentes valores de t . A figura põe em evidência que a solução da equação das ondas é uma combinação de duas ondas estacionárias, uma deslocando-se para a direita e a outra para a esquerda, ambas com velocidade c .

No conjunto de exercícios que se apresenta a seguir são dados outros exemplos de aplicação da generalização da regra de derivação das funções compostas no estudo das equações de derivadas parciais.

9.5. Exercícios

Nos exercícios que se seguem admite-se sempre a diferenciabilidade de todas as funções que se considerem.

1. Se k é uma constante positiva e $g(x, t) = \frac{1}{2} x \sqrt{kt}$, seja

$$f(x, t) = \int_0^{g(x, t)} e^{-u^2} du.$$

(a) Provar que $\frac{\partial f}{\partial x} = e^{-g^2} \frac{\partial g}{\partial x}$ e $\frac{\partial f}{\partial t} = e^{-g^2} \frac{\partial g}{\partial t}$.

- (b) Provar que f satisfaz à equação de derivadas parciais

$$k \frac{\partial^2 f}{\partial x^2} = \frac{\partial f}{\partial t} \quad (\text{a equação do calor})$$

2. Considerar um campo escalar f definido em \mathbb{R}^2 , tal que $f(x, y)$ dependa unicamente da distância r de (x, y) à origem, $f(x, y) = g(r)$, com $r = (x^2 + y^2)^{1/2}$.

- (a) Provar que para $(x, y) \neq (0, 0)$ se tem

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = \frac{1}{r} g'(r) + g''(r).$$

- (b) Supor além disso que f satisfaz à equação de Laplace

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0,$$

para todo $(x, y) \neq (0, 0)$. Demonstrar, recorrendo à alínea (a), que $f(x, y) = a \log(x^2 + y^2) + b$ para $(x, y) \neq (0, 0)$, onde a e b são constantes.

3. Repetir o Exercício 2 no caso n -dimensional, com $n \geq 3$, isto é, admitir que $f(x) = f(x_1, \dots, x_n) = g(r)$, com $r = \|x\|$. Mostrar que

$$\frac{\partial^2 f}{\partial x_1^2} + \dots + \frac{\partial^2 f}{\partial x_n^2} = \frac{n-1}{r} g'(r) + g''(r)$$

para $x \neq 0$. Se f satisfaz à equação de Laplace n -dimensional,

$$\frac{\partial^2 f}{\partial x_1^2} + \dots + \frac{\partial^2 f}{\partial x_n^2} = 0,$$

para todo $x \neq 0$, provar que $f(x) = a\|x\|^{2-n} + b$ para $x \neq 0$, com a e b constantes.

Nota: O operador linear ∇^2 definido pela equação

$$\nabla^2 f = \frac{\partial^2 f}{\partial x_1^2} + \cdots + \frac{\partial^2 f}{\partial x_n^2}$$

chama-se *laplaciano n -dimensional*.

4. *Laplaciano a duas dimensões em coordenadas polares.* A introdução de coordenadas polares $x = r \cos \theta$, $y = r \sin \theta$, transforma $f(x, y)$ em $g(r, \theta)$. Verificar as fórmulas seguintes:

$$(a) \quad \|\nabla f(r \cos \theta, r \sin \theta)\|^2 = \left(\frac{\partial g}{\partial r}\right)^2 + \frac{1}{r^2} \left(\frac{\partial g}{\partial \theta}\right)^2.$$

$$(b) \quad \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = \frac{\partial^2 g}{\partial r^2} + \frac{1}{r^2} \frac{\partial^2 g}{\partial \theta^2} + \frac{1}{r} \frac{\partial g}{\partial r}.$$

5. *Laplaciano a três dimensões em coordenadas esféricas.* A introdução de coordenadas esféricas

$$x = \rho \cos \theta \sin \varphi, \quad y = \rho \sin \theta \sin \varphi, \quad z = \rho \cos \varphi,$$

transforma $f(x, y, z)$ em $F(\rho, \theta, \varphi)$. Este exercício prova como proceder para exprimir o laplaciano $\nabla^2 f$ por intermédio das derivadas parciais de F .

(a) Introduzir em primeiro lugar as coordenadas polares $x = r \cos \theta$, $y = r \sin \theta$ para transformar $f(x, y, z)$ em $g(r, \theta, z)$. Recorrer ao Exercício 4 para provar que

$$\nabla^2 f = \frac{\partial^2 g}{\partial r^2} + \frac{1}{r^2} \frac{\partial^2 g}{\partial \theta^2} + \frac{1}{r} \frac{\partial g}{\partial r} + \frac{\partial^2 g}{\partial z^2}.$$

(b) Transformar $g(r, \theta, z)$ em $F(\rho, \theta, \varphi)$ fazendo $z = \rho \cos \varphi$, $r = \rho \sin \varphi$. Notar que, excepto para uma mudança de notação, esta transformação é a mesma que foi utilizada na alínea (a). Provar que

$$\nabla^2 f = \frac{\partial^2 F}{\partial \rho^2} + \frac{2}{\rho} \frac{\partial F}{\partial \rho} + \frac{1}{\rho^2} \frac{\partial^2 F}{\partial \varphi^2} + \frac{\cos \varphi}{\rho^2 \sin \varphi} \frac{\partial F}{\partial \varphi} + \frac{1}{\rho^2 \sin \varphi} \frac{\partial^2 F}{\partial \theta^2}.$$

6. Este exercício põe em evidência como aparece a equação de Legendre, quando se procuram soluções da equação de Laplace tendo uma forma especial. Seja f um campo escalar satisfazendo à equação de Laplace tridimensional, $\nabla^2 f = 0$. Introduzir coordenadas esféricas como no Exercício 5 e fazer $F(\rho, \theta, \varphi) = f(x, y, z)$.

(a) Admita-se que se procuram soluções da equação de Laplace tais que $F(\rho, \theta, \varphi)$ é independente de θ e tem a forma particular $F(\rho, \theta, \varphi) = \rho^n G(\varphi)$. Mostrar que f satisfaz à equação de Laplace se G satisfizer à equação diferencial ordinária de segunda ordem

$$\frac{d^2 G}{d\varphi^2} + \cot \varphi \frac{dG}{d\varphi} + n(n+1)G = 0.$$

(b) A mudança de variável $x = \cos \varphi$ ($\varphi = \arccos x$, $-1 \leq x \leq 1$) transforma $G(\varphi)$ em $g(x)$. Provar que g satisfaz à equação de Legendre

$$(1 - x^2) \frac{d^2 g}{dx^2} - 2x \frac{dg}{dx} + n(n+1)g = 0.$$

7. *Equação das ondas a duas dimensões.* Uma membrana delgada e flexível está estendida sobre o plano XOY e pode vibrar. Seja $z = f(x, y, t)$ o deslocamento da membrana, em relação a XOY , no ponto (x, y) e no instante t . Considerações físicas sugerem que f satisfaz à equação das ondas a duas dimensões,

$$\frac{\partial^2 f}{\partial t^2} = c^2 \left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \right),$$

sendo c uma constante positiva dependendo das características físicas da membrana. Este exercício revela uma conexão entre esta equação e a equação diferencial de Bessel.

- (a) Introduzir coordenadas polares $x = r \cos \theta$, $y = r \sin \theta$, e seja $F(r, \theta, t) = f(r \cos \theta, r \sin \theta, t)$. Se f verifica a equação das ondas, mostrar que F satisfaz à equação

$$\frac{\partial^2 F}{\partial t^2} = c^2 \left(\frac{\partial^2 F}{\partial r^2} + \frac{1}{r^2} \frac{\partial^2 F}{\partial \theta^2} + \frac{1}{r} \frac{\partial F}{\partial r} \right).$$

- (b) Se $F(r, \theta, t)$ é independente de θ , por exemplo $F(r, \theta, t) = \varphi(r, t)$ a equação em (a) simplifica-se para

$$\frac{\partial^2 \varphi}{\partial t^2} = c^2 \left(\frac{\partial^2 \varphi}{\partial r^2} + \frac{1}{r} \frac{\partial \varphi}{\partial r} \right).$$

Seja agora φ uma solução tal que $\varphi(r, t)$ se exprime pelo produto de uma função de r por uma função de t , $\varphi(r, t) = R(r)T(t)$. Provar que cada uma das funções R e T satisfaz a uma equação diferencial ordinária linear de segunda ordem.

- (c) Se a função T da alínea (b) é periódica com período $2\pi/c$, provar que R satisfaz à equação de Bessel $r^2 R'' + rR' + r^2 R = 0$.

9.6. Derivadas de funções implícitas

Certas superfícies no espaço tridimensional são representados por equações cartesianas da forma

$$F(x, y, z) = 0.$$

Uma tal equação diz-se que define numa *representação implícita* da superfície. Por exemplo, a equação $x^2 + y^2 + z^2 - 1 = 0$ representa a superfície de uma esfera unitária com centro na origem. Algumas vezes é possível resolver a equação $F(x, y, z) = 0$ em relação a uma das variáveis, exprimindo-a em função das duas restantes, por exemplo z em função de x e y . Isto conduz-nos a uma ou mais equações da forma

$$z = f(x, y).$$

Para a esfera obtermos duas soluções,

$$z = \sqrt{1 - x^2 - y^2} \quad \text{e} \quad z = -\sqrt{1 - x^2 - y^2},$$

uma representando a semi-esfera superior, a outra a semi-esfera inferior.

No caso geral pode não ser uma questão fácil obter uma fórmula explícita para z em função de x e y . Por exemplo, não existe nenhum método simples que permita resolver, em ordem a z , a equação $y^2 + xz + z^2 - e^z - 4 = 0$. Contudo, uma utilização adequada da regra de derivação da função composta torna possível deduzir várias propriedades das derivadas $\partial f / \partial x$ e $\partial f / \partial y$ sem um conhecimento explícito de $f(x, y)$. O processo é analisado nesta secção.

Admitamos que existe uma função $f(x, y)$ tal que

$$(9.19) \quad F[x, y, f(x, y)] = 0$$

para todo (x, y) em algum conjunto aberto S , embora não seja possível obter fórmulas explícitas para o cálculo de $f(x, y)$. Expressimos tal circunstância dizendo que a equação $F(x, y, z) = 0$ define z *implicitamente* como uma função de x e y e escrevemos

$$z = f(x, y).$$

Introduzamos agora uma função auxiliar g definida em S do modo seguinte:

$$g(x, y) = F[x, y, f(x, y)].$$

A equação (9.19) estabelece que $g(x, y) = 0$ em S , logo as derivadas parciais $\partial g / \partial x$ e $\partial g / \partial y$ são também 0 em S . Mas podemos também calcular estas derivadas parciais pela regra da derivação de uma função composta. Para o conseguir escrevemos

$$g(x, y) = F[u_1(x, y), u_2(x, y), u_3(x, y)],$$

onde $u_1(x, y) = x$, $u_2(x, y) = y$, e $u_3(x, y) = f(x, y)$. A regra referida dá-nos as fórmulas

$$\frac{\partial g}{\partial x} = D_1 F \frac{\partial u_1}{\partial x} + D_2 F \frac{\partial u_2}{\partial x} + D_3 F \frac{\partial u_3}{\partial x} \quad \text{e} \quad \frac{\partial g}{\partial y} = D_1 F \frac{\partial u_1}{\partial y} + D_2 F \frac{\partial u_2}{\partial y} + D_3 F \frac{\partial u_3}{\partial y},$$

onde cada derivada parcial $D_k F$ deve ser calculada em $(x, y, f(x, y))$. Uma vez que se tem

$$\frac{\partial u_1}{\partial x} = 1, \quad \frac{\partial u_2}{\partial x} = 0, \quad \frac{\partial u_3}{\partial x} = \frac{\partial f}{\partial x}, \quad \text{e} \quad \frac{\partial g}{\partial x} = 0,$$

a primeira das equações precedentes escreve-se

$$D_1 F + D_3 F \frac{\partial f}{\partial x} = 0.$$

Resolvendo esta em ordem a $\partial f / \partial x$ obtemos

$$(9.20) \quad \frac{\partial f}{\partial x} = - \frac{D_1 F[x, y, f(x, y)]}{D_3 F[x, y, f(x, y)]}$$

naqueles pontos para os quais $D_3 F[x, y, f(x, y)] \neq 0$. Por um raciocínio análogo obtemos a fórmula correspondente para $\partial f / \partial y$:

$$(9.21) \quad \frac{\partial f}{\partial y} = - \frac{D_2 F[x, y, f(x, y)]}{D_3 F[x, y, f(x, y)]}$$

nos pontos para os quais $D_3 F[x, y, f(x, y)] \neq 0$. Estas fórmulas escrevem-se habitualmente na forma mais sintética:

$$\frac{\partial f}{\partial x} = - \frac{\partial F / \partial x}{\partial F / \partial z}, \quad \frac{\partial f}{\partial y} = - \frac{\partial F / \partial y}{\partial F / \partial z}.$$

EXEMPLO. Admitamos que a equação $y^2 + xz + z^2 - e^z - c = 0$ define z como uma função x e y , por exemplo $z = f(x, y)$. Determinar um valor para a constante c tal que $f(0, e) = 2$, e calcular as derivadas parciais $\partial f / \partial x$ e $\partial f / \partial y$ no ponto $(x, y) = (0, e)$.

Resolução. Quando $x = 0$, $y = e$, e $z = 2$, a equação escreve-se $e^2 + 4 - e^2 - c = 0$, a qual indica que $c = 4$. Seja $F(x, y, z) = y^2 + xz + z^2 - e^z - 4$. De (9.20) e (9.21) obtém-se

$$\frac{\partial f}{\partial x} = - \frac{z}{x + 2z - e^z}, \quad \frac{\partial f}{\partial y} = - \frac{2y}{x + 2z - e^z}.$$

Quando $x = 0$, $y = e$, e $z = 2$ encontramos $\partial f / \partial x = 2/(e^2 - 4)$ e $\partial f / \partial y = 2e/(e^2 - 4)$. Observe-se que é possível calcular as derivadas parciais $\partial f / \partial x$ e $\partial f / \partial y$ recorrendo unicamente ao valor de $f(x, y)$ no ponto $(0, e)$.

A discussão acabada de fazer pode ser generalizada a funções de mais do que duas variáveis.

TEOREMA 9.3. *Seja F um campo escalar diferenciável num conjunto aberto T de \mathbb{R}^n . Supondo que a equação*

$$F(x_1, \dots, x_n) = 0$$

define x_n implicitamente como uma função diferenciável de x_1, \dots, x_{n-1} ,

$$x_n = f(x_1, \dots, x_{n-1}),$$

para todos os pontos (x_1, \dots, x_{n-1}) em algum conjunto aberto S de \mathbb{R}^{n-1} , então para cada $k = 1, 2, \dots, n-1$, a derivada parcial $D_k f$ é dada pela fórmula

$$(9.22) \quad D_k f = - \frac{D_k F}{D_n F}$$

nos pontos em que $D_n F \neq 0$. As derivadas parciais $D_k F$ e $D_n F$ que aparecem em (9.22) devem calcular-se no ponto $(x_1, x_2, \dots, x_{n-1}, f(x_1, x_2, \dots, x_{n-1}))$.

A demonstração constitui uma generalização directa do raciocínio efectuado para deduzir as equações (9.20) e (9.21) e deixa-se ao cuidado do leitor.

A discussão pode generalizar-se noutro sentido. Admitamos que temos duas superfícies com as seguintes representações implícitas:

$$(9.23) \quad F(x, y, z) = 0, \quad G(x, y, z) = 0.$$

Se estas superfícies se intersectam ao longo de uma curva C , pode ser possível obter numa representação paramétrica de C resolvendo as duas equações (9.23) simultaneamente em relação a duas das variáveis em função da terceira, por exemplo x e y em função de z . Suponhamos que é possível resolvê-la relativamente a x e y e que as soluções sejam dadas pelas equações

$$x = X(z), \quad y = Y(z)$$

para todo z em certo intervalo aberto (a, b) . Então quando x e y se substituem por $X(z)$ e $Y(z)$, respectivamente, as duas equações em (9.23) são identicamente satisfeitas, isto é, podemos escrever $F[X(z), Y(z), z] = 0$ e $G[X(z), Y(z), z] = 0$ para todo z em (a, b) . Mais uma vez, recorrendo à regra de derivação da função composta, podemos calcular as derivadas $X'(z)$ e $Y'(z)$ sem um conhecimento explícito de $X(z)$ e $Y(z)$. Com esse objectivo introduzamos novas funções f e g definidas por

$$f(z) = F[X(z), Y(z), z] \quad \text{e} \quad g(z) = G[X(z), Y(z), z].$$

Então $f(z) = g(z) = 0$ para todo z em (a, b) e consequentemente as derivadas $f'(z)$ e $g'(z)$ são também nulas em (a, b) . Pela regra de derivação referida estas derivadas são definidas pelas fórmulas

$$f'(z) = \frac{\partial F}{\partial x} X'(z) + \frac{\partial F}{\partial y} Y'(z) + \frac{\partial F}{\partial z}, \quad g'(z) = \frac{\partial G}{\partial x} X'(z) + \frac{\partial G}{\partial y} Y'(z) + \frac{\partial G}{\partial z}.$$

Visto que $f'(z)$ e $g'(z)$ são ambas nulas, podemos determinar $X'(z)$ e $Y'(z)$ pela resolução do seguinte par de equações lineares simultâneas:

$$\begin{aligned} \frac{\partial F}{\partial x} X'(z) + \frac{\partial F}{\partial y} Y'(z) &= - \frac{\partial F}{\partial z}, \\ \frac{\partial G}{\partial x} X'(z) + \frac{\partial G}{\partial y} Y'(z) &= - \frac{\partial G}{\partial z}. \end{aligned}$$

Nos pontos em que o determinante do sistema não for nulo, estas equações têm uma solução única a qual, de acordo com a regra de Cramer, pode exprimir-se do modo seguinte:

$$(9.24) \quad X'(z) = - \frac{\begin{vmatrix} \frac{\partial F}{\partial z} & \frac{\partial F}{\partial y} \\ \frac{\partial G}{\partial z} & \frac{\partial G}{\partial y} \end{vmatrix}}{\begin{vmatrix} \frac{\partial F}{\partial x} & \frac{\partial F}{\partial y} \\ \frac{\partial G}{\partial x} & \frac{\partial G}{\partial y} \end{vmatrix}}, \quad Y'(z) = - \frac{\begin{vmatrix} \frac{\partial F}{\partial x} & \frac{\partial F}{\partial z} \\ \frac{\partial G}{\partial x} & \frac{\partial G}{\partial z} \end{vmatrix}}{\begin{vmatrix} \frac{\partial F}{\partial x} & \frac{\partial F}{\partial y} \\ \frac{\partial G}{\partial x} & \frac{\partial G}{\partial y} \end{vmatrix}}.$$

Os determinantes figurando em (9.24) são determinantes de matrizes jacobianas e dizem-se, por isso, *determinantes jacobianos*. Uma notação especial é por vezes usada para representar este tipo de determinantes. Escrever-se

$$\frac{\partial(f_1, \dots, f_n)}{\partial(x_1, \dots, x_n)} = \det \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \dots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \dots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}.$$

Recorrendo a esta notação, as fórmulas (9.24) podem exprimir-se mais resumidamente na forma

$$(9.25) \quad X'(z) = \frac{\partial(F, G)/\partial(y, z)}{\partial(F, G)/\partial(x, y)}, \quad Y'(z) = \frac{\partial(F, G)/\partial(z, x)}{\partial(F, G)/\partial(x, y)}.$$

(O sinal menos incorporou-se nos numeradores por troca das colunas).

O método pode ser generalizado para tratar situações mais gerais nas quais são dadas m equações com n variáveis, sendo $n > m$, obtendo-se m variáveis em função das $n - m$ restantes. As derivadas parciais das novas funções assim definidas podem exprimir-se como cocientes de determinantes jacobianos, generalizando (9.25).

Um exemplo com $m = 2$ e $n = 4$ apresenta-se no Exercício 3 da Secção 9.8.

9.7. Exemplos resolvidos

Vamos apresentar alguns dos conceitos da anterior secção na resolução de alguns problemas relativos a funções definidas implicitamente.

EXEMPLO 1. Supor que a equação $g(x, y) = 0$ define y como uma função derivável de x , seja $y = Y(x)$, para todo x em algum intervalo aberto (a, b) . Expressar a derivada $Y'(x)$ em função das derivadas parciais de g .

Resolução. Seja $G(x) = g[x, Y(x)]$ para x em (a, b) . Então a equação $g(x, y) = 0$ implica $G(x) = 0$ em (a, b) . Pela regra de derivação da função composta temos

$$G'(x) = \frac{\partial g}{\partial x} \cdot 1 + \frac{\partial g}{\partial y} Y'(x),$$

donde se obtém

$$(9.26) \quad Y'(x) = - \frac{\partial g / \partial x}{\partial g / \partial y}$$

nos pontos x de (a, b) para os quais $\partial g / \partial y \neq 0$. As derivadas parciais $\partial g / \partial x$ e $\partial g / \partial y$ são dadas pelas fórmulas $\partial g / \partial x = D_1 g[x, Y(x)]$ e $\partial g / \partial y = D_2 g[x, Y(x)]$.

EXEMPLO 2. Quando se elimina y das duas equações $z = f(x, y)$ e $g(x, y) = 0$, o resultado pode exprimir-se na forma $z = h(x)$. Escrever a derivada $h'(x)$ em função das derivadas parciais de f e g .

Solução. Admitamos que a equação $g(x, y) = 0$ pode resolver-se relativamente a y em função de x e que a solução é dada por $y = Y(x)$ para todo x em certo intervalo aberto (a, b) . Então a função h é dada pela fórmula

$$h(x) = f[x, Y(x)] \quad \text{se } x \in (a, b).$$

Aplicando a regra de derivação da função composta temos

$$h'(x) = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} Y'(x).$$

Resolvendo a equação (9.26) do Exemplo 1 obtemos a fórmula

$$h'(x) = \frac{\frac{\partial g}{\partial y} \frac{\partial f}{\partial x} - \frac{\partial f}{\partial y} \frac{\partial g}{\partial x}}{\frac{\partial g}{\partial y}}.$$

As derivadas parciais do segundo membro são calculadas no ponto $(x, Y(x))$. Observe-se que o numerador pode também exprimir-se como um determinante jacobiano, dando-nos

$$h'(x) = \frac{\partial(f, g) / \partial(x, y)}{\partial g / \partial y}.$$

EXEMPLO 3. As duas equações $2x = v^2 - u^2$ e $y = uv$ definem u e v como funções de x e y . Determinar fórmulas correspondentes a $\partial u/\partial x$, $\partial u/\partial y$, $\partial v/\partial x$, $\partial v/\partial y$.

Resolução. Mantendo y fixo e derivando ambas as equações relativamente a x , tendo presente que u e v são funções de x e y , obtemos

$$2 = 2v \frac{\partial v}{\partial x} - 2u \frac{\partial u}{\partial x} \quad \text{e} \quad 0 = u \frac{\partial v}{\partial x} + v \frac{\partial u}{\partial x}.$$

Resolvendo estas equações simultâneas em relação a $\partial v/\partial x$ e $\partial u/\partial x$ obtemos

$$\frac{\partial u}{\partial x} = -\frac{u}{u^2 + v^2} \quad \text{e} \quad \frac{\partial v}{\partial x} = \frac{v}{u^2 + v^2}.$$

Por outro lado, se fixamos x e derivamos ambas as equações relativamente a y obtemos

$$0 = 2v \frac{\partial v}{\partial y} - 2u \frac{\partial u}{\partial y} \quad \text{e} \quad 1 = u \frac{\partial v}{\partial y} + v \frac{\partial u}{\partial y}.$$

Para solução deste sistema de equações encontramos

$$\frac{\partial u}{\partial y} = \frac{v}{u^2 + v^2} \quad \text{e} \quad \frac{\partial v}{\partial y} = \frac{u}{u^2 + v^2}.$$

EXEMPLO 4. Seja n definido como função de x e y pela equação

$$u = F(x + u, yu).$$

Determinar $\partial u/\partial x$ e $\partial u/\partial y$ por intermédio das derivadas parciais de F .

Resolução. Suponhamos que $u = g(x, y)$ para todo (x, y) em algum conjunto aberto S . Substituindo $g(x, y)$ por u na equação original obtemos

$$(9.27) \quad g(x, y) = F[u_1(x, y), u_2(x, y)],$$

onde $u_1(x, y) = x + g(x, y)$ e $u_2(x, y) = y g(x, y)$. Mantenhamos agora y fixo e derivemos ambos os membros de (9.27) relativamente a x , aplicando no segundo membro a regra de derivação da função composta, obtendo

$$(9.28) \quad \frac{\partial g}{\partial x} = D_1 F \frac{\partial u_1}{\partial x} + D_2 F \frac{\partial u_2}{\partial x}.$$

Mas $\partial u_1/\partial x = 1 + \partial g/\partial x$, e $\partial u_2/\partial x = y \partial g/\partial x$. Por conseguinte (9.28) escreve-se

$$\frac{\partial g}{\partial x} = D_1 F \cdot \left(1 + \frac{\partial g}{\partial x}\right) + D_2 F \cdot \left(y \frac{\partial g}{\partial x}\right).$$

Resolvendo esta equação relativamente a $\partial g/\partial x$ (e escrevendo $\partial u/\partial x$ em vez de $\partial g/\partial x$) obtemos

$$\frac{\partial u}{\partial x} = \frac{-D_1 F}{D_1 F + y D_2 F - 1}.$$

Do mesmo modo encontramos

$$\frac{\partial g}{\partial y} = D_1 F \frac{\partial u_1}{\partial y} + D_2 F \frac{\partial u_2}{\partial y} = D_1 F \frac{\partial g}{\partial y} + D_2 F \left(y \frac{\partial g}{\partial y} + g(x, y)\right).$$

Isto conduz-nos à equação

$$\frac{\partial u}{\partial y} = \frac{-g(x, y) D_2 F}{D_1 F + y D_2 F - 1}.$$

As derivadas parciais $D_1 F$ e $D_2 F$ são calculadas no ponto $(x + g(x, y), yg(x, y))$.

EXEMPLO 5. Quando se elimina u entre as equações $x = u + v$ e $y = uv^2$, obtemos uma equação da forma $F(x, y, v) = 0$, a qual define v implicitamente como uma função de x e y , por exemplo $v = h(x, y)$. Provar que

$$\frac{\partial h}{\partial x} = \frac{h(x, y)}{3h(x, y) - 2x}$$

e determinar uma fórmula semelhante para $\partial h/\partial y$.

Resolução. Eliminando u entre as duas equações obtém-se

$$xv^2 - v^3 - y = 0.$$

Seja F a função definida por

$$F(x, y, v) = xv^2 - v^3 - y.$$

A discussão da Secção 9.6 é aplicável aqui e podemos escrever

$$(9.29) \quad \frac{\partial h}{\partial x} = -\frac{\partial F/\partial x}{\partial F/\partial v} \quad \text{e} \quad \frac{\partial h}{\partial y} = -\frac{\partial F/\partial y}{\partial F/\partial v}.$$

Mas $\partial F/\partial x = v^2$, $\partial F/\partial v = 2xv - 3v^2$, e $\partial F/\partial y = -1$. Por conseguinte, a equação (9.29) escreve-se

$$\frac{\partial h}{\partial x} = -\frac{v^2}{2xv - 3v^2} = -\frac{v}{2x - 3v} = \frac{h(x, y)}{3h(x, y) - 2x}$$

$$\frac{\partial h}{\partial y} = -\frac{-1}{2xv - 3v^2} = \frac{1}{2xh(x, y) - 3h^2(x, y)}.$$

EXEMPLO 6. A equação $F(x, y, z) = 0$ define z implicitamente como uma função de x e y , seja $z = f(x, y)$. Supondo que $\partial^2 F / (\partial x \partial z) = \partial^2 F / (\partial z \partial x)$, provar que

$$(9.30) \quad \frac{\partial^2 f}{\partial x^2} = -\frac{\left(\frac{\partial^2 F}{\partial z^2}\right)\left(\frac{\partial F}{\partial x}\right)^2 - 2\left(\frac{\partial^2 F}{\partial x \partial z}\right)\left(\frac{\partial F}{\partial z}\right)\left(\frac{\partial F}{\partial x}\right) + \left(\frac{\partial F}{\partial z}\right)^2\left(\frac{\partial^2 F}{\partial x^2}\right)}{\left(\frac{\partial F}{\partial z}\right)^3},$$

onde as derivadas parciais do segundo membro são calculadas em $(x, y, f(x, y))$.

Resolução. Pela equação (9.20) da Secção 9.6 temos

$$(9.31) \quad \frac{\partial f}{\partial x} = -\frac{\partial F / \partial x}{\partial F / \partial z}.$$

Devemos ter presente que na realidade este cociente significa

$$-\frac{D_1 F[x, y, f(x, y)]}{D_3 F[x, y, f(x, y)]}.$$

Introduzamos $G(x, y) = D_1 F[x, y, f(x, y)]$ e $H(x, y) = D_3 F[x, y, f(x, y)]$. É nosso propósito calcular a derivada parcial, com respeito a x , do cociente

$$\frac{\partial f}{\partial x} = -\frac{G(x, y)}{H(x, y)},$$

mantendo y fixo. A regra de derivação do cociente dá-nos

$$(9.32) \quad \frac{\partial^2 f}{\partial x^2} = -\frac{H \frac{\partial G}{\partial x} - G \frac{\partial H}{\partial x}}{H^2}.$$

Uma vez que G e H são funções compostas, usamos a correspondente regra de derivação para calcular as derivadas parciais $\partial G / \partial x$ e $\partial H / \partial x$. Para $\partial G / \partial x$ temos

$$\begin{aligned} \frac{\partial G}{\partial x} &= D_1(D_1 F) \cdot 1 + D_2(D_1 F) \cdot 0 + D_3(D_1 F) \cdot \frac{\partial f}{\partial x} \\ &= \frac{\partial^2 F}{\partial x^2} + \frac{\partial^2 F}{\partial z \partial x} \frac{\partial f}{\partial x}. \end{aligned}$$

Analogamente, encontramos

$$\begin{aligned}\frac{\partial H}{\partial x} &= D_1(D_3 F) \cdot 1 + D_2(D_3 F) \cdot 0 + D_3(D_3 F) \cdot \frac{\partial f}{\partial x} \\ &= \frac{\partial^2 F}{\partial x \partial z} + \frac{\partial^2 F}{\partial z^2} \frac{\partial f}{\partial x}.\end{aligned}$$

Introduzindo estas derivadas em (9.32) e substituindo $\partial f / \partial x$ pelo segundo membro de (9.31) obtemos a fórmula (9.30).

9.8. Exercícios

Nos exercícios apresentados a seguir deve admitir-se sempre a existência e continuidade das derivadas que se considerem.

1. As duas equações $x + y = uv$ e $xy = u - v$ determinam x e y implicitamente como funções de u e v , seja $x = X(u, v)$ e $y = Y(u, v)$. Mostrar que $\partial X / \partial u = (xv - 1)/(x - y)$ se $x \neq y$ e determinar fórmulas análogas para $\partial X / \partial v$, $\partial Y / \partial u$, $\partial Y / \partial v$.
2. As duas equações $x + y = uv$ e $xy = u - v$ determinam x e v como funções de u e y , a saber $x = X(u, y)$ e $v = V(u, y)$. Mostrar que $\partial X / \partial u = (u + v)/(1 + yu)$ se $1 + yu \neq 0$, e determinar fórmula análogas para $\partial X / \partial y$, $\partial V / \partial u$, $\partial V / \partial y$.
3. As duas equações $F(x, y, u, v) = 0$ e $G(x, y, u, v) = 0$ determinam x e y implicitamente como funções de u e v , sejam $x = X(u, v)$ e $y = Y(u, v)$. Mostrar que

$$\frac{\partial X}{\partial u} = \frac{\partial(F, G) / \partial(y, u)}{\partial(F, G) / \partial(x, y)}$$

em pontos para os quais o jacobiano $\partial(F, G) / \partial(x, z) \neq 0$, e determinar fórmulas semelhantes para as derivadas parciais $\partial X / \partial v$, $\partial Y / \partial u$, e $\partial Y / \partial v$.

4. A intersecção de duas superfícies definidas pelas equações cartesianas $2x^2 + 3y^2 - z^2 = 25$ e $x^2 + y^2 = z^2$ define uma curva C passando pelo ponto $P = (\sqrt{7}, 3, 4)$. Estas equações podem resolver-se em relação a x e a y em função de z para darem uma representação paramétrica de C , com z como parâmetro.
 - (a) Determinar o vector tangente unitária T a C no ponto P sem recorrer ao conhecimento explícito da representação paramétrica.
 - (b) Confrontar o resultado da alínea (a) com o obtido mediante uma representação paramétrica de C com z como parâmetro.
5. As três equações $F(u, v) = 0$, $u = xy$, e $v = \sqrt{x^2 + z^2}$ definem uma superfície no espaço $OXYZ$. Determinar o vector normal a esta superfície no ponto $x = 1$, $y = 1$, $z = \sqrt{3}$ se for sabido que $D_1 F(1, 2) = 1$ e $D_2 F(1, 2) = 2$.
6. As três equações

$$x^2 - y \cos(uv) + z^2 = 0,$$

$$x^2 + y^2 - \sin(uv) + 2z^2 = 2,$$

$$xy - \sin u \cos v + z = 0,$$

definem x , y e z como funções de u e v . Calcular as derivadas parciais $\partial x/\partial u$ e $\partial x/\partial v$ no ponto $x = y = 1$, $u = \pi/2$, $v = 0$, $z = 0$.

7. A equação $f(y/x, z/x) = 0$ define z implicitamente como uma função de x e y , seja $z = g(x, y)$. Mostrar que

$$x \frac{\partial g}{\partial x} + y \frac{\partial g}{\partial y} = g(x, y)$$

nos pontos em que $D_2 f[y/x, g(x, y)/x]$ é diferente de zero.

8. Seja F uma função real de duas variáveis reais e suponhamos que as derivadas parciais $d_1 F$ e $D_2 F$ são diferentes de zero. Seja u outra função real de duas variáveis reais tal que as derivadas parciais $\partial u/\partial x$ e $\partial u/\partial y$ estejam relacionadas pela equação $F(\partial u/\partial x, \partial u/\partial y) = 0$. Provar que existe uma constante n tal que

$$\frac{\partial^2 u}{\partial x^2} \frac{\partial^2 u}{\partial y^2} = \left(\frac{\partial^2 u}{\partial x \partial y} \right)^n,$$

e determinar n . Supor que $\partial^2 u/(\partial x \partial y) = \partial^2 u/(\partial y \partial x)$.

9. A equação $x + z + (y + z)^2 = 6$ define z como função implícita de x e y , $z = f(x, y)$. Calcular as derivadas parciais $\partial f/\partial x$, $\partial f/\partial y$ e $\partial^2 f/(\partial x \partial y)$ em função de x , y e z .
10. A equação $\sin(x + y) + \sin(y + z) = 1$ define z como função implícita de x e y , $z = f(x, y)$. Calcular a derivada $D_1 z$ em função de x , y e z .
11. A equação $F(x + y + z, x^2 + y^2 + z^2) = 0$ define z como função implícita de x e y , $z = f(x, y)$. Determinar as derivadas parciais $\partial f/\partial x$ e $\partial f/\partial y$ em função de $D_1 F$ e $D_2 F$.
12. Seja f e g duas funções de uma variável real e definamos $F(x, y) = f[x + g(y)]$. Achar as fórmulas correspondentes a todas as derivadas parciais de F de primeira e segunda ordem, expressas em função das derivadas de f e g . Verificar a relação

$$\frac{\partial F}{\partial x} \frac{\partial^2 F}{\partial x \partial y} = \frac{\partial F}{\partial y} \frac{\partial^2 F}{\partial x^2}.$$

9.9. Máximos, mínimos e pontos sela

Uma superfície, definida explicitamente por uma equação da forma $z = f(x, y)$, pode ser considerada como uma superfície de nível do campo escalar F , definida pela equação

$$F(x, y, z) = f(x, y) - z.$$

Se f é diferenciável, a gradiente deste campo é definido pelo vector

$$\nabla F = \frac{\partial f}{\partial x} \mathbf{i} + \frac{\partial f}{\partial y} \mathbf{j} - \mathbf{k}.$$

A equação linear que define o plano tangente no ponto $P_1 = (x_1, y_1, z_1)$ pode escrever-se

$$z - z_1 = A(x - x_1) + B(y - y_1),$$

com

$$A = D_1 f(x_1, y_1) \quad \text{e} \quad B = D_2 f(x_1, y_1).$$

Quando ambos os coeficientes A e B são nulos, o ponto P_1 diz-se um *ponto de estacionaridade* da superfície o ponto (x_1, y_1) diz-se um *ponto de estacionaridade* ou um *ponto crítico* da função f . O plano tangente à superfície, num tal ponto, é paralelo a XOY . Os pontos de estacionaridade de uma superfície classificam-se habitualmente em três categorias: máximos, mínimos e pontos sela. Se a superfície se imaginar como um terreno montanhoso, essas categorias correspondem, respectivamente ao cume da montanha, ao fundo dos vales, e às gargantas.

O conceito de máximo, mínimo e pontos sela pode definir-se para campos escalares arbitrários, definidos em subconjuntos de \mathbb{R}^n .

DEFINIÇÃO. Diz-se que um campo escalar f tem um máximo absoluto num ponto a de um conjunto S de \mathbb{R}^n se

$$(9.33) \quad f(x) \leq f(a)$$

para todo x em S . O mínimo $f(a)$ chama-se o valor máximo absoluto de f em S . A função f diz-se ter um máximo relativo em a se a desigualdade em (9.33) é satisfeita para todo o x pertencente a uma certa n -bola $B(a)$ de S .

Por outras palavras, um máximo relativo em a é o máximo absoluto em certa vizinhança de a . Os *mínimo absoluto* e *mínimo relativo* definem-se de uma maneira análoga, recorrendo a uma desigualdade de sentido contrário à de (9.33). Os adjectivos global e local são algumas vezes utilizados em vez dos termos *absoluto* e *relativo*, respectivamente.

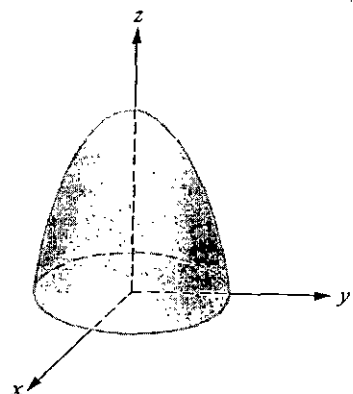
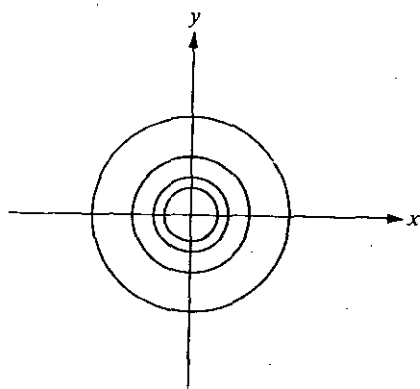
DEFINIÇÃO. Um número que seja quer máximo relativo ou mínimo relativo de f chama-se em extremo de f .

Se f possui um extremo num ponto interior a e é aí diferenciável, então todas as derivadas parciais de primeira ordem $D_1 f(a), \dots, D_n f(a)$ devem anular-se: Por outras palavras, $\nabla f(a) = \mathbf{0}$. (Isto pode demonstrar-se facilmente mantendo fixa cada componente e reduzindo o problema ao caso unidimensional.) No caso em que $n = 2$, significa o que acabou de referir-se que existe um plano tangente à superfície $z = f(x, y)$ no ponto $(a, f(a))$ paralelo ao plano XOY . Por outro lado, é fácil construir exemplos nos quais o anulamento de todas as derivadas parciais em a não implica necessariamente um extremo em a . Verifica-se tal circunstância nos chamados pontos sela que se definem a seguir.

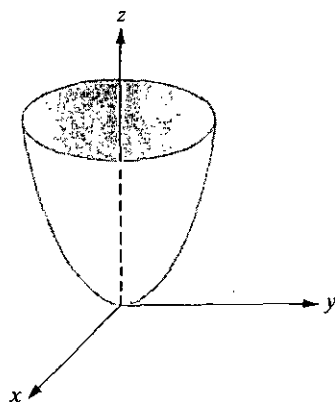
DEFINIÇÃO. Se f é diferenciável em a e se $\nabla f(a) = \mathbf{0}$, o ponto a diz-se de estacionaridade de f . Um ponto de estacionaridade de f diz-se um ponto sela se toda a n -bola $B(a)$ contém pontos x tais que $f(x) < f(a)$ e outros pontos para os quais $f(x) > f(a)$.

A situação é algo parecida com o caso unidimensional em que os pontos de estacionaridade se classificam em máximos, mínimos e pontos de inflexão. Os exemplos que se apresentam a seguir ilustram estes diferentes tipos de pontos de estacionaridade. Em todos os exemplos os pontos de estacionaridade em questão situam-se na origem.

EXEMPLO 1. Máximo relativo. $z = f(x, y) = 2 - x^2 - y^2$. Tal superfície é um parabolóide de revolução. Na vizinhança da origem tem a forma representada na figura 9.3(a). As curvas de nível são círculos, alguns dos quais se encontram desenhados na figura 9.3(b). Uma vez que $f(x, y) = 2 - (x^2 + y^2) < 2 = f(0, 0)$ para todo o par (x, y) ,

(a) $z = 2 - x^2 - y^2$ (b) Curvas de nível: $x^2 + y^2 = c$.

Exemplo 1: Máximo relativo na origem.

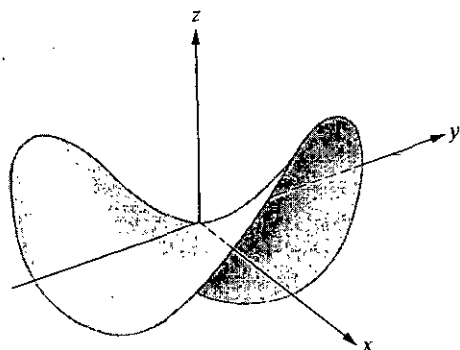
(c) $z = x^2 + y^2$

Exemplo 2: Mínimo relativo na origem.

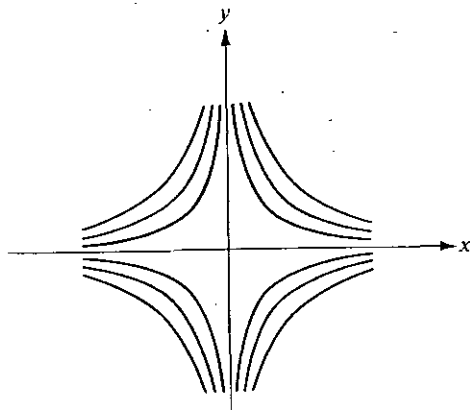
FIG. 9.3. Exemplos 1 e 2.

resulta que f não só tem um máximo relativo em $(0, 0)$, mas ele é também um máximo absoluto. Ambas as derivadas parciais $\partial f / \partial x$ e $\partial f / \partial y$ se anulam na origem.

EXEMPLO 2. Mínimo relativo. $z = f(x, y) = x^2 + y^2$. Esta superfície, outro parabolóide de revolução, é essencialmente a mesma que a do Exemplo 1, excepto em que existe um mínimo na origem em vez de um máximo. A forma da superfície, na vizinhança da origem, está representada na figura 9.3(c) e algumas das suas curvas de nível estão traçadas em 9.3(b).



(a) $z = xy$

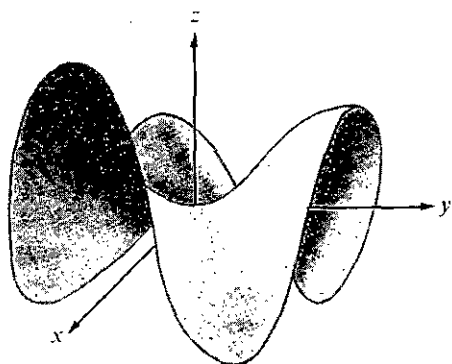


(b) Curvas de nível: $x^3 - 3xy^2 = c$.

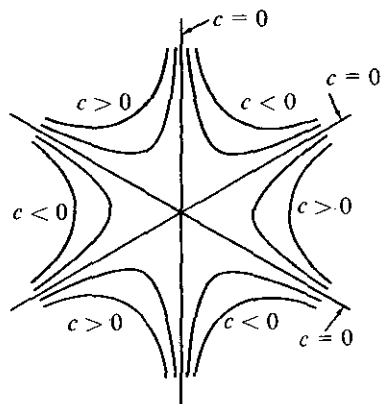
FIG. 9.4. Exemplo 3. Ponto sela na origem.

EXEMPLO 3. Ponto sela. $z = f(x, y) = xy$. A superfície, é um parabolóide hiperbólico. Na vizinhança da origem é semelhante a uma sela, como se representa na fig. 9.4(a). Ambas as derivadas parciais $\partial f / \partial x$ e $\partial f / \partial y$ são nulas na origem, mas não existe nem máximo nem mínimo relativo nesse ponto. Con efeito, para pontos (x, y) no primeiro e terceiro quadrantes, x e y têm o mesmo sinal, donde resulta $f(x, y) > 0 = f(0, 0)$, enquanto que para pontos nos segundo e quarto quadrantes x e y tem sinais opostos, dando-nos $f(x, y) < 0 = f(0, 0)$. Consequentemente, em cada vizinhança da origem existem pontos nos quais a função é menor que $f(0, 0)$ e pontos em que a função excede $f(0, 0)$, pelo que a origem é um ponto sela. A presença de tal ponto é revelada igualmente na figura 9.4(b), na qual se representam algumas das curvas de nível nas vizinhanças de $(0, 0)$. Estas são hipérbolas, admitindo os eixos coordenados como assíntotas.

EXEMPLO 4. Ponto sela. $z = f(x, y) = x^3 - 3xy^2$. Na proximidade da origem, esta superfície tem o aspecto dum colo duma montanha na vizinhança de três picos. Esta superfície está representada na figura 9.5(a), estando traçadas na figura 9.5(b) algumas curvas de nível. É evidente que a origem é um ponto sela.

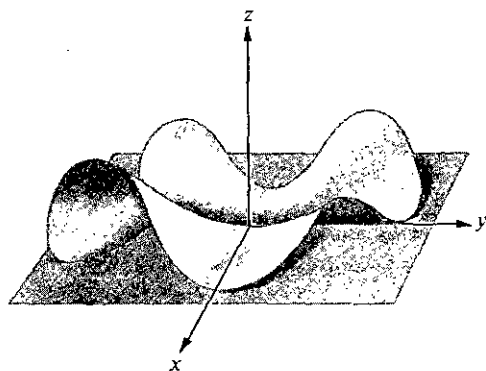


(a) $z = x^3 - 3xy^2$.

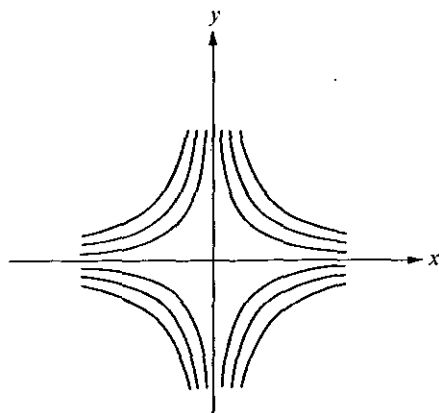


(b) Curvas de nível: $x^3 - 3xy^2 = c$.

FIG. 9.5. Exemplo 4. Ponto sela na origem.

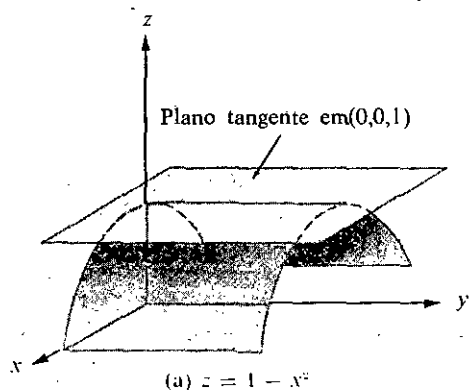


(a) $z = x^2y^2$.

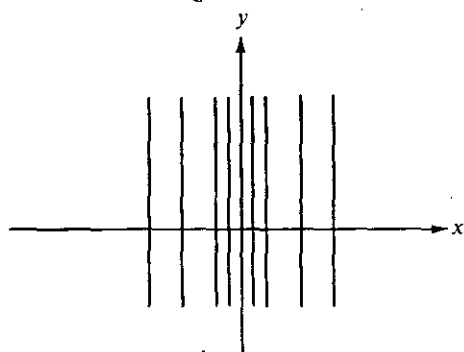


(b) Curvas de nível: $x^2y^2 = c$.

FIG. 9.6. Exemplo 5. Mínimo relativo na origem.



(a) $z = 1 - x^2$.



(b) Curvas de nível: $1 - x^2 = c$.

FIG. 9.7. Exemplo 6. Máximo relativo na origem.

EXEMPLO 5. Mínimo relativo. $z = f(x, y) = x^2 y^2$. Esta superfície tem o aspecto de um vale circundado por quatro montanhas, como sugere a figura 9.6(a). Existe um mínimo absoluto na origem, uma vez que $f(x, y) \geq f(0, 0)$ para todo (x, y) . As curvas de nível [representadas na figura 9.6(b)] são hipérbolas admitindo os eixos coordenados por assíntotas. Observe-se que estas curvas de nível são semelhantes às dos Exemplo 3. Neste caso, porém, a função toma unicamente valores não negativos ao longo de todas as suas curvas de nível.

EXEMPLO 6. Máximo relativo. $z = f(x, y) = 1 - x^2$. Neste caso a superfície é um cilindro de geratrizes paralelas ao eixo OY , como se mostra na figura 9.7(a). As secções planas, definidas por planos paralelos ao eixo OX , são parábolas. Existe obviamente um máximo absoluto na origem, devido ao facto de $f(x, y) = 1 - x^2 \leq 1 = f(0, 0)$ para todo o ponto (x, y) . As curvas de nível constituem uma família de rectas paralelas como se indica na figura 9.7(b).

9.10. Fórmula de Taylor de segunda ordem para campos escalares

Se um campo escalar diferenciável f admite um ponto de estacionaridade em \mathbf{a} , a natureza de tal ponto fica determinada pelo sinal da diferença $f(\mathbf{x}) - f(\mathbf{a})$ para \mathbf{x} próximo de \mathbf{a} . Se $\mathbf{x} = \mathbf{a} + \mathbf{y}$, então temos a fórmula de Taylor de primeira ordem

$$f(\mathbf{a} + \mathbf{y}) - f(\mathbf{a}) = \nabla f(\mathbf{a}) \cdot \mathbf{y} + \|\mathbf{y}\| E(\mathbf{a}, \mathbf{y}), \text{ onde } E(\mathbf{a}, \mathbf{y}) \rightarrow 0 \text{ quando } \mathbf{y} \rightarrow \mathbf{0}.$$

Num ponto de estacionaridade, $\nabla f(\mathbf{a}) = \mathbf{0}$ e a fórmula de Taylor vem

$$f(\mathbf{a} + \mathbf{y}) - f(\mathbf{a}) = \|\mathbf{y}\|^2 E(\mathbf{a}, \mathbf{y}).$$

Para determinar o sinal algébrico de $f(\mathbf{a} + \mathbf{y}) - f(\mathbf{a})$ necessitamos mais informação relativa ao termo de erro $\|\mathbf{y}\|^2 E(\mathbf{a}, \mathbf{y})$. O teorema que se segue prova que se f admite derivadas parciais de segunda ordem contínuas em \mathbf{a} , o termo $\|\mathbf{y}\|^2 E(\mathbf{a}, \mathbf{y})$ é igual à forma quadrática

$$\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n D_{ij} f(\mathbf{a}) y_i y_j$$

mais um termo de ordem inferior a $\|\mathbf{y}\|^2$. Os coeficientes da forma quadrática são as derivadas parciais de segunda ordem $D_{ij} f = D_i(D_j f)$, calculadas em \mathbf{a} . A matriz, $n \times n$, das derivadas parciais de segunda ordem $D_{ij} f(\mathbf{x})$ chama-se a *matriz hessiana*† e representa-se por $H(\mathbf{x})$. Então, temos

† Devida a Ludwig Otto Hesse (1811-1874), um matemático alemão autor de muitas contribuições à teoria das superfícies.

$$H(x) = [D_{ij}f(x)]_{i,j=1}^n$$

sempre que tais derivadas existam. A forma quadrática pode escrever-se mais simplesmente na notação matricial como segue:

$$\sum_{i=1}^n \sum_{j=1}^n D_{ij}f(a) y_i y_j = yH(a)y^t,$$

onde $y = (y_1, \dots, y_n)$ se considera como uma matriz linha $1 \times n$ e y^t é a sua transposta, uma matriz coluna $n \times 1$. Quando as derivadas parciais $D_{ij}f$ são contínuas, temos $D_{ij}f = D_{ji}f$ e a matriz $H(a)$ é simétrica.

A fórmula de Taylor, dando uma aproximação quadrática para $f(a+y) - f(a)$, toma agora a seguinte forma.

TEOREMA 9.4. FÓRMULA DE TAYLOR DE SEGUNDA ORDEM PARA CAMPOS ESCALARES. *Se f é um campo escalar, admitindo derivadas parciais de segunda ordem, $D_{ij}f$, contínuas numa n -bola $B(a)$, então para todo y em \mathbb{R}^n tal que $a+y \in B(a)$ tem-se*

$$(9.34) \quad f(a+y) - f(a) = \nabla f(a) \cdot y + \frac{1}{2!} yH(a+cy)y^t, \quad \text{com} \quad 0 < c < 1.$$

Esta pode também escrever-se na forma

$$(9.35) \quad f(a+y) - f(a) = \nabla f(a) \cdot y + \frac{1}{2!} yH(a)y^t + \|y\|^2 E_2(a, y),$$

onde $E_2(a, y) \rightarrow 0$ quando $y \rightarrow 0$.

Demonstração. Consideremos y fixo e definamos $g(u)$, para u real, pela igualdade

$$g(u) = f(a + uy) \quad \text{para} \quad -1 \leq u \leq 1.$$

Então $f(a+y) - f(a) = g(1) - g(0)$. Demonstraremos o teorema aplicando a fórmula de Taylor de segunda ordem a g no intervalo $[0, 1]$. Obtemos

$$(9.36) \quad g(1) - g(0) = g'(0) + \frac{1}{2!} g''(c), \quad \text{com} \quad 0 < c < 1.$$

Aqui utilizou-se para o resto a fórmula de Lagrange (ver Secção 7.7 do Volume I).

Visto ser g uma função composta definida por $g(u) = f(r(u))$, onde $r(u) = a + uy$, podemos calcular a sua derivada aplicando a correspondente regra. Temos $r'(u) = y$, pelo que a regra de derivação da função composta nos conduz a

$$g'(u) = \nabla f[r(u)] \cdot r'(u) = \nabla f[r(u)] \cdot y = \sum_{j=1}^n D_{ij}f[r(u)] y_j,$$

desde que $r(u) \in B(a)$. Em particular, $g'(0) = \Delta f(a) \cdot y$. Aplicando mais uma vez a mesma regra de derivação encontramos

$$g''(u) = \sum_{i=1}^n D_i \left(\sum_{j=1}^n D_j f[r(u)] y_j \right) y_i = \sum_{i=1}^n \sum_{j=1}^n D_{ij} f[r(u)] y_i y_j = yH[r(u)]y^t.$$

Deste modo $g''(c) = yH(a + cy)y^t$, pelo que a equação (9.36) se transforma na (9.34).

Para demonstrar (9.35) definimos $E_2(a, y)$ pela equação

$$(9.37) \quad \|y\|^2 E_2(a, y) = \frac{1}{2!} y \{H(a + cy) - H(a)\} y^t \quad \text{se } y \neq 0,$$

e seja $E_2(a, 0) = 0$. Então a equação (9.34) toma a forma

$$f(a + y) - f(a) = \nabla f(a) \cdot y + \frac{1}{2!} yH(a)y^t + \|y\|^2 E_2(a, y).$$

Para completar a demonstração necessitamos mostrar que $E_2(a, y) \rightarrow 0$ quando $y \rightarrow 0$.

De (9.37), obtemos

$$\begin{aligned} \|y\|^2 |E_2(a, y)| &= \frac{1}{2} \left| \sum_{i=1}^n \sum_{j=1}^n \{D_{ij}f(a + cy) - D_{ij}f(a)\} y_i y_j \right| \\ &\leq \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n |D_{ij}f(a + cy) - D_{ij}f(a)| \|y\|^2. \end{aligned}$$

Dividindo por $\|y\|^2$ obtemos a desigualdade

$$|E_2(a, y)| \leq \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n |D_{ij}f(a + cy) - D_{ij}f(a)|$$

para $y \neq 0$. Porque cada derivada parcial de segunda ordem $D_{ij}f$ é contínua em a , temos $D_{ij}f(a + cy) \rightarrow D_{ij}f(a)$ quando $y \rightarrow 0$, de modo que $E_2(a, y) \rightarrow 0$ quando $y \rightarrow 0$, o que completa a demonstração.

9.11. A natureza do ponto de estacionaridade determinada pelos valores próprios da matriz Hessiana

Num ponto de estacionaridade tem-se $\nabla f(a) = 0$, pelo que a fórmula de Taylor dada em (9.35) vem

$$f(a + y) - f(a) = \frac{1}{2} yH(a)y^t + \|y\|^2 E_2(a, y).$$

Visto o termo $\|y\|^2 E_2(a, y)$ tender para zero mais rapidamente que $\|y\|^2$, parece razoável esperar que para valores de y pequenos o sinal de $f(a + y) - f(a)$ seja o mesmo da forma quadrática $yH(a)y^t$; consequentemente a natureza de um ponto de estacionaridade poderá ser determinado pelo sinal da forma quadrática. Esta Secção é dedicada à demonstração desta afirmação.

Em primeiro lugar estabelecemos uma conexão entre o sinal da forma quadrática e os seus valores próprios.

TEOREMA 9.5. *Seja $A = [a_{ij}]$ uma matriz simétrica real $n \times n$ e considere-se*

$$Q(y) = yAy^t = \sum_{i=1}^n \sum_{j=1}^n a_{ij} y_i y_j.$$

tem-se então:

- (a) $Q(y) > 0$ para todo $y \neq 0$ se e só se todos os valores próprios de A são positivos.
- (b) $Q(y) < 0$ para todo $y \neq 0$ se e só se todos os valores próprios de A são negativos.

Nota: Na hipótese (a), a forma quadrática diz-se *definida positiva*; na hipótese (b) diz-se *definida negativa*.

Demonstração. Segundo o teorema 5.11 existe uma matriz ortogonal C que reduz a forma quadrática yAy^t à forma diagonal, isto é,

$$(9.38) \quad Q(y) = yAy^t = \sum_{i=1}^n \lambda_i x_i^2$$

onde $x = (x_1, \dots, x_n)$ é a matriz coluna $x = yC$, e $\lambda_1, \dots, \lambda_n$ são os valores próprios de A . Os valores próprios são reais visto que A é simétrica.

Se todos os valores próprios são positivos, a equação (9.38) mostra que $Q(y) > 0$ sempre que $x \neq 0$. Mas porque $x = yC$ temos $y = xC^{-1}$, de maneira que $x \neq 0$ se e só se $y \neq 0$. Portanto $Q(y) > 0$ para todo $y \neq 0$.

Inversamente, se $Q(y) > 0$ para todo $y \neq 0$ podemos escolher y de maneira que $x = yC$ seja o vector coordenado e_k . Para este y , a equação (9.38) dá-nos $Q(y) = \lambda_k$, de maneira que cada $\lambda_k > 0$. Está assim demonstrada a alínea (a). A demonstração de (b) é inteiramente análoga.

O teorema apresentado a seguir descreve a natureza de pontos de estacionaridade em função do sinal da forma quadrática $yH(a)y^t$.

TEOREMA 9.6. *Se f é um campo escalar com derivadas parciais de segunda ordem $D_{ij}f$ contínuas numa n -bola $B(a)$, e $H(a)$ representa a matriz hessiana num ponto de estacionaridade a , então tem-se:*

- (a) *Se todos os valores próprios de $H(a)$ são positivos, f tem um mínimo relativo em a .*
- (b) *Se todos os valores próprios de $H(a)$ são negativos, f tem um máximo relativo em a .*
- (c) *Se $H(a)$ tem valores próprios positivos e negativos, então f tem um ponto sela em a .*

Demonstração. Seja $Q(y) = yH(a)y^t$. A fórmula de Taylor dá-nos

$$(9.39) \quad f(a + y) - f(a) = \frac{1}{2}Q(y) + \|y\|^2 E_2(a, y),$$

onde $E_2(a, y) \rightarrow 0$ quando $y \rightarrow 0$. Provaremos que existe um número positivo r tal que, se $0 < \|y\| < r$, o sinal algébrico de $f(a + y) - f(a)$ é o mesmo que o de $Q(y)$.

Admitamos em primeiro lugar que todos os valores próprios $\lambda_1, \dots, \lambda_n$ de $H(a)$ são positivos. Seja h o menor valor próprio. Se $u < h$, os n números

$$\lambda_1 - u, \dots, \lambda_n - u$$

são também positivos. Estes números são os valores próprios da matriz simétrica real $H(a) - uI$, onde I é a matriz identidade $n \times n$. Pelo teorema 9.5, a forma quadrática $y[H(a) - uI]y^t$ é definida positiva, e consequentemente $y[H(a) - uI]y^t > 0$ para todo $y \neq O$. Portanto

$$yH(a)y^t > y(uI)y^t = u \|y\|^2$$

para todo real $u < h$. Tomando $u = \frac{1}{2}h$ obtemos a desigualdade

$$Q(y) > \frac{1}{2}h \|y\|^2$$

para todo $y \neq O$. Porque $E_2(a, y) \rightarrow O$ quando $y \rightarrow O$, existe um número positivo r tal que $|E_2(a, y)| < \frac{1}{4}h$ sempre que $0 < \|y\| < r$. Para tal y temos

$$0 \leq \|y\|^2 |E_2(a, y)| < \frac{1}{4}h \|y\|^2 < \frac{1}{2}Q(y),$$

e a fórmula de Taylor (9.39) mostra que

$$f(a + y) - f(a) \geq \frac{1}{2}Q(y) - \|y\|^2 |E_2(a, y)| > 0.$$

Deste modo f tem um mínimo relativo em a , o que demonstra a alínea (a). Para demonstrar (b) podemos seguir um raciocínio análogo, ou mais simplesmente aplicar a alínea (a) a $-f$.

Para demonstrar (c), sejam λ_1 , e λ_2 dois valores próprios de $H(a)$ de sinais contrários. Seja $h = \min \{|\lambda_1|, |\lambda_2|\}$. Então para cada real u satisfazendo a $-h < u < h$ os números

$$\lambda_1 - u \quad \text{e} \quad \lambda_2 - u$$

são valores próprios de sinais contrários para a matriz $H(a) - uI$. Portanto, se $u \in (-h, h)$, a forma quadrática $y[H(a) - uI]y^t$ torna quer valores positivos quer negativos em toda a vizinhança de $y = O$. Escolhamos, como atrás, um $r > 0$ tal que $|E_2(a, y)| < \frac{1}{4}h$ sempre que $0 < \|y\| < r$. Então, argumentando do mesmo modo, vemos que para tal y o sinal de $f(a + y) - f(a)$ é o mesmo de $Q(y)$. Uma vez que quando $y \rightarrow O$ aparecem valores positivos e negativos, f tem em a um ponto sela. Esta conclui a completa demonstração do teorema.

Nota: Se todos os valores próprios de $H(a)$ são nulos, o teorema 9.6 não dá qualquer informação relativa ao ponto de estacionaridade. Podem estabelecer-se critérios para estudar tais exemplos, os quais fazem intervir derivadas de ordem superior; estes porém não serão tratados aqui.

9.12. Critério das derivadas de segunda ordem para extremos de funções de duas variáveis

No caso em que $n=2$ a natureza de um ponto de estacionaridade pode também determinar-se pelo da derivada de segunda ordem $D_{1,1}f(a)$ e do determinante da matriz Hessiana.

TEOREMA 9.7. *Seja a um ponto de estacionaridade de um campo escalar $f(x_1, x_2)$ admitindo derivadas parciais de segunda ordem contínuas numa 2-bola $B(a)$ e sejam*

$$A = D_{1,1}f(a), \quad B = D_{1,2}f(a), \quad C = D_{2,2}f(a),$$

e

$$\Delta = \det H(a) = \det \begin{bmatrix} A & B \\ B & C \end{bmatrix} = AC - B^2.$$

Então verifica-se:

- (a) Se $\Delta < 0$, f admite um ponto sela em a .
- (b) Se $\Delta > 0$ e $A > 0$, f admite um mínimo relativo em a .
- (c) Se $\Delta > 0$ e $A < 0$, f admite um máximo relativo em a .
- (d) Se $\Delta = 0$, o critério não é conclusivo.

Demonstração. Neste caso a equação característica $\det [\lambda I - H(a)] = 0$ é uma equação quadrática,

$$\lambda^2 - (A + C)\lambda + \Delta = 0.$$

Os valores próprios λ_1, λ_2 estão relacionados com os coeficientes pelas equações

$$\lambda_1 + \lambda_2 = A + C, \quad \lambda_1 \lambda_2 = \Delta.$$

Se $\Delta < 0$ os valores próprios tem sinais contrários, pelo que f admite um ponto sela em a , o que demonstra (a). Se $\Delta > 0$, os valores próprios têm o mesmo sinal. Neste caso $AC > B^2 \geq 0$, pelo que A e C têm o mesmo sinal. Este sinal terá forçosamente que ser o de λ_1 e λ_2 , pois $A + C = \lambda_1 + \lambda_2$, o que prova (b) e (c).

Para demonstrar (d) referir-nos-emos aos Exemplos 4 e 5 da Secção 9.9. Em ambos os exemplos temos $\Delta = 0$ na origem. No Exemplo 4 a origem é um ponto sela, e no Exemplo 5 é um mínimo relativo.

Ainda o teorema 9.7 seja aplicável, pode não constituir a maneira mais simples de determinar a natureza de um ponto de estacionaridade. Por exemplo, quando $f(x, y) = e^{1/g(x, y)}$ com $g(x, y) = x^2 + 2 + \cos^2 y - 2 \cos y$, o critério é aplicável, mas os cálculos são extensos. Neste caso podemos exprimir $g(x, y)$ como uma soma escrevendo $g(x, y) = 1 + x^2 + (1 - \cos y)^2$. Vemos por sua vez que f tem máximos relativos nos pontos para os quais $x^2 = 0$ e $(1 - \cos y)^2 = 0$. Estes são os pontos $(0, 2n\pi)$, quando n é qualquer inteiro.

9.13. Exercícios

Nos Exercícios 1 a 15, determinar e classificar (caso existam) os pontos de estacionaridade das superfícies cujas equações cartesianas se indicam

1. $z = x^2 + (y - 1)^2$.
2. $z = x^2 - (y - 1)^2$.
3. $z = 1 + x^2 - y^2$.
4. $z = (x - y + 1)^2$.
5. $z = 2x^2 - xy - 3y^2 - 3x + 7y$.
6. $z = x^2 - xy + y^2 - 2x + y$.
7. $z = x^3 - 3xy^2 + y^3$.
8. $z = x^2y^2(6 - x - y)$.
9. $z = x^3 + y^3 - 3xy$.
10. $z = \sin x \cosh y$.
11. $z = e^{2x+3y}(8x^2 - 6xy + 3y^2)$.
12. $z = (5x + 7y - 25)e^{-(x^2+xy+y^2)}$.
13. $z = \sin x \sin y \sin(x + y)$, $0 \leq x \leq \pi$, $0 \leq y \leq \pi$.
14. $z = x - 2y + \log \sqrt{x^2 + y^2} + 3 \arctan \frac{y}{x}$, $x > 0$.
15. $z = (x^2 + y^2)e^{-(x^2+y^2)}$.
16. Seja $f(x, y) = 3x^4 - 4x^2y + y^2$. Mostrar que sobre toda a recta $y = mx$ a função tem um mínimo em $(0, 0)$, mas que não existe mínimo relativo em nenhuma vizinhança bidimensional da origem. Fazer um desenho indicando o conjunto de pontos (x, y) para os quais $f(x, y) > 0$ e o conjunto para os quais $f(x, y) < 0$.
17. Seja $f(x, y) = (3-x)(3-y)(x+y-3)$.
 - (a) Fazer um desenho representando o conjunto de pontos (x, y) para os quais $f(x, y) \geq 0$.
 - (b) Determinar todos os pontos (x, y) no plano para os quais $D_1f(x, y) = D_2f(x, y) = 0$. [Sugestão: $D_1f(x, y)$ admite $(3-y)$ como factor.]
 - (c) Quais dos pontos de estacionaridade são máximos relativos? Quais são mínimos relativos? Quais são nem mínimos nem máximos? Justificar as respostas.
 - (d) Terá f um mínimo absoluto ou um máximo absoluto em todo o plano? Justificar as respostas.
18. Determinar todos os valores extremos, relativos e absolutos, e os pontos sela para a função $f(x, y) = xy(1 - x^2 - y^2)$ no quadro $0 \leq x \leq 1$, $0 \leq y \leq 1$.
19. Determinar as constantes a e b tais que o integral

$$\int_0^1 \{ax + b - f(x)\}^2 dx$$

seja o menor possível se (a) $f(x) = x^2$; (b) $f(x) = (x^2 + 1)^{-1}$.

20. Seja $f(x, y) = Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F$, com $A > 0$ e $B^2 < AC$.
 - (a) Provar que existe um ponto (x_1, y_1) no qual f tem um mínimo. [Sugestão: Transformar a parte quadrática numa soma de quadrados.]
 - (b) Provar que $f(x_1, y_1) = Dx_1 + Ey_1 + F$ nesse mínimo.
 - (c) Mostrar que

$$f(x_1, y_1) = \frac{1}{AC - B^2} \begin{vmatrix} A & B & D \\ B & C & E \\ D & E & F \end{vmatrix}$$

21. **Método dos menores quadrados.** Dados n números distintos x_1, \dots, x_n e n outros números y_1, \dots, y_n (não necessariamente distintos), é em geral impossível determinar uma recta $f(x) = ax + b$ que passe pelos pontos (x_i, y_i) , isto é, tal que $f(x_i) = y_i$ para cada i . Contudo podemos encontrar uma função linear a qual torna mínimo o "erro quadrático total"

$$E(a, b) = \sum_{i=1}^n [f(x_i) - y_i]^2$$

Determinar os valores de a e b para os quais isso se verifica.

22. Generalizar o método dos menores quadrados ao espaço tridimensional, isto é, determinar uma função linear $f(x, y) = ax + by + c$ a qual minimize o erro quadrático total

$$E(a, b, c) = \sum_{i=1}^n [f(x_i, y_i) - z_i]^2,$$

onde (x_i, y_i) são n pontos distintos dados e z_1, \dots, z_n são n números reais dados.

23. Sejam z_1, \dots, z_n n pontos distintos num m espaço. Se $x \in \mathbb{R}^m$, definamos

$$f(x) = \sum_{k=1}^n \|x - z_k\|^2.$$

Provar que f tem um mínimo no ponto $a = \frac{1}{n} \sum_{k=1}^n z_k$ (centroide).

24. Seja a um ponto de estacionaridade de um campo escalar f , com derivadas parciais de segunda ordem contínuas numa n -bola $B(a)$. Provar que f admite um ponto sela em a se pelo menos dois dos elementos diagonais da matriz Hessiana $H(a)$ têm sinais contrários.
25. Verificar que o campo escalar $f(x, y, z) = x^4 + y^4 + z^4 - 4xyz$ admite um ponto de estacionaridade em $(1, 1, 1)$, e determinar a natureza desse ponto de estacionaridade pelo cálculo dos valores próprios da sua matriz Hessiana.

9.14. Extremos condicionados. Multiplicadores de Lagrange

Iniciamos esta Secção com dois exemplos de problemas de extremos sujeitos a determinadas condições (ou ligações).

EXEMPLO 1. Dada uma superfície S que não contém a origem, determinar os pontos de S mais da origem.

EXEMPLO 2. Se $f(x, y, z)$ representa a temperatura em (x, y, z) , determinar os valores máximo e mínimo da temperatura ao longo de uma dada curva C no espaço tridimensional.

Ambos os exemplos são casos particulares do seguinte problema geral: *Determinar os valores de um campo escalar $f(x)$ quando x está sujeito à restrição de pertencer a um dado subconjunto do domínio de f .*

No Exemplo 1 o campo escalar a ser minimizado é a função distância,

$$f(x, y, z) = (x^2 + y^2 + z^2)^{1/2};$$

o subconjunto restrictivo é a superfície dada S . No Exemplo 2 o subconjunto é a curva dada C .

Os problemas de extremos condicionados são, na maior parte das vezes, muito complicados; não se conhece um método geral que permita resolvê-los em toda a sua generalidade. Métodos particulares são possíveis quando o subconjunto restrictivo tem uma estrutura simples, por exemplo, se for uma superfície como no Exemplo 1, ou uma curva como no Exemplo 2. Nesta Secção discutimos o método dos multiplicadores de Lagrange aplicável na resolução de tais problemas. Vamos primeiramente descrever o método na sua forma geral, e depois apresentaremos argumentos geométricos para mostrarmos como é aplicável aos dois exemplos atrás mencionados.

O método dos multiplicadores de Lagrange. Se um campo escalar $f(x_1, \dots, x_n)$ admite um extremo relativo quando sujeito a m condições, por exemplo

$$(9.40) \quad g_1(x_1, \dots, x_n) = 0, \quad \dots, \quad g_m(x_1, \dots, x_n) = 0,$$

onde $m < n$, então existem m escalares $\lambda_1, \dots, \lambda_m$ tais que

$$(9.41) \quad \nabla f = \lambda_1 \nabla g_1 + \dots + \lambda_m \nabla g_m$$

em cada ponto extremo.

Na prática para determinarmos os pontos extremos consideramos o sistema de $n + m$ equações formadas com as m equações de condição (9.40) e as n equações escalares determinadas pela relação vectorial (9.41). Tais equações devem resolver-se (se possível) em relação às $n + m$ incógnitas x_1, \dots, x_n e $\lambda_1, \dots, \lambda_m$. Os pontos (x_1, \dots, x_n) em que ocorrem extremos relativos encontram-se entre as soluções dessas equações.

Os escalares $\lambda_1, \dots, \lambda_m$ introduzidos como auxiliares na resolução deste tipo de problemas chamam-se os *multiplicadores de Lagrange*. Introduz-se um multiplicador para cada condição. O campo escalar f e as funções de condição g_1, \dots, g_m supõem-se diferenciáveis. O método é válido se o número de condições, m , é menor do que o número de variáveis n , e se nem todos os determinantes jacobianos das funções de condição com respeito a m das variáveis x_1, \dots, x_n são nulos no valor extremo em questão. A demonstração da validade do método é um resultado importante estudado em cursos mais avançados de Cálculo e não será discutido aqui (Ver Cap. 7 da obra do autor *Mathematical Analysis*, Addison-Wesley, tradução espanhola *Análisis Matemático*, Editorial Reverté, S. A. Barcelona). Em vez disso apresentaremos argumentos geométricos para mostrarmos como o método é aplicável nos dois exemplos enunciados no início desta Secção.

Resolução Geométrica do Exemplo 1. Pretendemos determinar os pontos de uma dada superfície S que estão mais próximos da origem. Um ponto (x, y, z) do espaço tridimensional dista r da origem se e só se estiver sobre a esfera

$$x^2 + y^2 + z^2 = r^2.$$

Esta esfera é uma superfície de nível da função $f(x, y, z) = (x^2 + y^2 + z^2)^{1/2}$ a qual há que minimizar. Se começamos com $r = 0$ e aumentarmos r até que a referida superfície de nível seja tangente à superfície dada S , cada ponto de contacto será um ponto de S o mais próximo possível da origem. Para determinarmos as coordenadas dos pontos de contacto admitimos que S é definida por uma equação cartesiana da forma $g(x, y, z) = 0$. Se S admite plano tangente em cada ponto de contacto, este plano deverá ser igualmente tangente à superfície de nível tangente a S nesse mesmo ponto. Portanto o vector gradiente de superfície $g(x, y, z) = 0$ deve ser paralelo ao vector gradiente da superfície de nível de contacto, $f(x, y, z) = r$. Deste modo existirá uma constante λ tal que

$$\nabla f(x, y, z) = \lambda \nabla g(x, y, z)$$

em cada ponto de contacto. Esta é a equação vectorial (9.41) obtida pelo método de Lagrange quando existir uma única condição.

Resolução geométrica do Exemplo 2. Pretendemos obter os valores extremos de uma função temperatura $f(x, y, z)$ ao longo de uma dada curva C . Se consideramos a curva C como a intersecção de duas superfícies, por exemplo

$$g_1(x, y, z) = 0 \quad \text{e} \quad g_2(x, y, z) = 0,$$

estamos perante um problema de extremos com duas condições. Os dois vectores gradiente ∇g_1 e ∇g_2 são normais a essas superfícies, logo também o são a C , a curva de intersecção. (Ver figura 9.8.) Mostraremos a seguir que o vector gradiente ∇f da função temperatura é também normal a C em cada extremo relativo sobre C . Isto

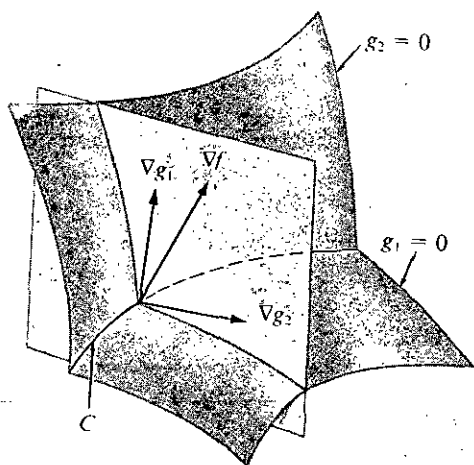


FIG. 9.8. Os vectores ∇g_1 , ∇g_2 e ∇f estão situados no mesmo plano.

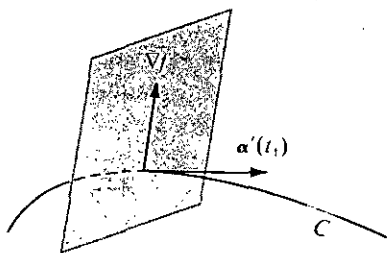


FIG. 9.9. O vector gradiente ∇f está situado num plano normal a C .

implica que ∇f está situado no mesmo plano que ∇g_1 e ∇g_2 ; consequentemente se ∇g_1 e ∇g_2 são independentes podemos exprimir ∇f como uma combinação linear de ∇g_1 e ∇g_2 , seja

$$\nabla f = \lambda_1 \nabla g_1 + \lambda_2 \nabla g_2.$$

Esta é a equação vectorial (9.41) proporcionada pelo método de Lagrange quando existem duas condições.

Para provar que ∇f é normal a C num ponto extremo, imaginamos C como sendo definida por uma função vectorial $\alpha(t)$, onde t varia num dado intervalo $[a, b]$. Sobre a curva C a temperatura vem numa função de t , por exemplo $\varphi(t) = f[\alpha(t)]$. Se φ admite um extremo relativo num ponto interior t_1 de $[a, b]$ devemos ter $\varphi'(t_1) = 0$. Por outro lado, a regra de derivação da função composta diz-nos que $\varphi'(t)$ é dada pelo produto escalar

$$\varphi'(t) = \nabla f[\alpha(t)] \cdot \alpha'(t).$$

Este produto escalar é nulo em t_1 , logo ∇f é perpendicular a $\alpha'(t_1)$. Mas $\alpha'(t_1)$ é tangente a C , pelo que $\nabla f[\alpha(t_1)]$ está num plano normal a C , como se indica na figura 9.9.

Os dois vectores gradiente ∇g_1 e ∇g_2 são independentes se e só se o seu produto vectorial é não nulo. O produto vectorial é dado por

$$\nabla g_1 \times \nabla g_2 = \begin{vmatrix} i & j & k \\ \frac{\partial g_1}{\partial x} & \frac{\partial g_1}{\partial y} & \frac{\partial g_1}{\partial z} \\ \frac{\partial g_2}{\partial x} & \frac{\partial g_2}{\partial y} & \frac{\partial g_2}{\partial z} \end{vmatrix} = \frac{\partial(g_1, g_2)}{\partial(y, z)} i + \frac{\partial(g_1, g_2)}{\partial(z, x)} j + \frac{\partial(g_1, g_2)}{\partial(x, y)} k.$$

Portanto, a independência de ∇g_1 e ∇g_2 significa que nenhum dos três determinantes jacobianos do segundo membro são nulos. Como foi referido anteriormente, o método de Lagrange é aplicável sempre que esta condição seja satisfeita.

Se ∇g_1 e ∇g_2 são dependentes o método pode falhar. Por exemplo, suponhamos que tentarmos aplicar o método de Lagrange para a determinação dos valores extremos de $f(x, y, z) = x^2 + y^2$ ao longo da curva de intersecção de duas superfícies $g_1(x, y, z) = 0$ e $g_2(x, y, z) = 0$, onde $g_1(x, y, z) = z$ e $g_2(x, y, z) = z^2 - (y - 1)^2$. As duas superfícies, um plano e um cilindro, intersectam-se segundo a recta C representada na figura 9.10. O problema obviamente tem uma solução, porque $f(x, y, z)$ representa a distância do ponto (x, y, z) ao eixo OZ e esta distância é mínima em C quando o ponto é $(0, 1, 0)$. Todavia, nesse ponto os vectores gradiente são $\nabla g_1 = k$, $\nabla g_2 = 0$, e $\nabla f = 2j$, e é evidente que não existem escalares λ_1 e λ_2 que verifiquem a equação (9.41).

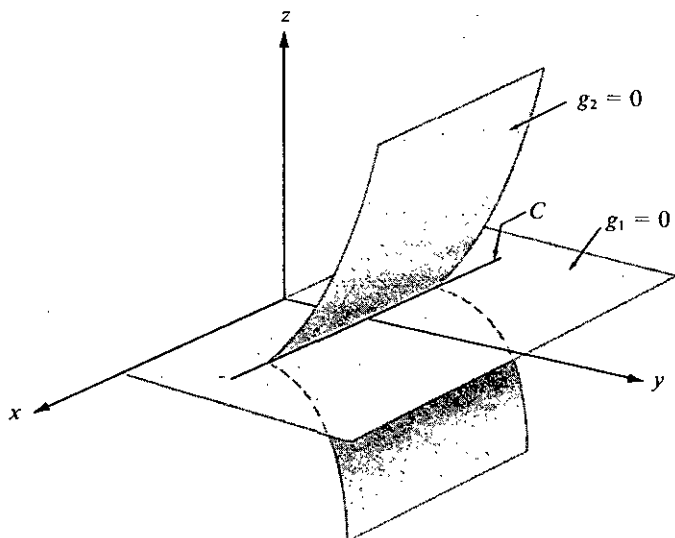


FIG. 9.10. Um exemplo em que o método de Lagrange não é aplicável.

9.15. Exercícios

1. Determinar os valores extremos de $z = xy$, com a condição $x + y = 1$.
2. Determinar as distâncias máxima e mínima da origem à curva $5x^2 + 6xy + 5y^2 = 8$.
3. Supor a e b números positivos fixos.
 - (a) Determinar os valores extremos de $z = x/a + y/b$ sob a condição de $x^2 + y^2 = 1$.
 - (b) Determinar os valores extremos de $z = x^2 + y^2$ sob a condição de $x/a + y/b = 1$.
 Para cada caso, interpretar geometricamente os problemas.
4. Determinar os extremos de $z = \cos^2 x + \cos^2 y$, sob a condição $x - y = \pi/4$.
5. Determinar os valores extremos do campo escalar $f(x, y, z) = x - 2y + 2z$ sobre a esfera $x^2 + y^2 + z^2 = 1$.
6. Determinar os pontos de $z^2 - xy = 1$ mais próximos da origem.
7. Determinar a menor distância do ponto $(1, 0)$ à parábola $y^2 = 4x$.
8. Determinar os pontos da curva de intersecção de duas superfícies

$$x^2 - xy + y^2 - z^2 = 1 \quad \text{e} \quad x^2 + y^2 = 1$$

que estão o mais próximo possível da origem.

9. Se a , b e c são números positivos, determinar o máximo valor de $f(x, y, z) = x^a y^b z^c$ sob a condição de $x + y + z = 1$.
10. Determinar o volume máximo limitado pelos planos $x = 0$, $y = 0$, $z = 0$, e um plano que é tangente ao elipsoide

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

um ponto do octante $x > 0$, $y > 0$, $z > 0$.

11. Determinar o máximo de $\log x + \log y + 3 \log z$ sobre a parte da esfera $x^2 + y^2 + z^2 = 5r^2$ em que $x > 0, y > 0, z > 0$. Utilizar o resultado para provar que para números reais e positivos a, b, c se tem

$$abc^3 \leq 27 \left(\frac{a+b+c}{5} \right)^5.$$

12. Dada a secção cónica $Ax^2 + 2Bxy + Cy^2 = 1$, em que $A > 0$ e $B^2 < AC$, representem m e M as distâncias da origem aos pontos mais próximos e mais afastado da cónica. Provar que

$$M^2 = \frac{A + C + \sqrt{(A - C)^2 + 4B^2}}{2(AC - B^2)}$$

e determinar uma fórmula análoga para m^2 .

13. Utilizar o método dos multiplicadores de Lagrange para determinar a máxima e mínima distâncias de um ponto da elipse $x^2 + 4y^2 = 4$ à recta $x + y = 4$.
14. A secção de um canal é um trapézio isósceles. Se os lados iguais do trapézio medem c metros, qual deve ser o ângulo que formam esses lados com o fundo (base menor do trapézio) se quisermos que a área da secção seja máxima?

9.16. Teorema do valor extremo para campos escalares contínuos

O teorema do valor extremo para funções contínuas reais num intervalo fechado e limitado pode generalizar-se a campos escalares. Consideremos campos escalares contínuos num intervalo fechado n -dimensional. Tal intervalo define-se como o produto cartesiano de n intervalos fechados unidimensionais. Se $\mathbf{a} = (a_1, \dots, a_n)$ e $\mathbf{b} = (b_1, \dots, b_n)$ escrevemos

$$[\mathbf{a}, \mathbf{b}] = [a_1, b_1] \times \dots \times [a_n, b_n] = \{(x_1, \dots, x_n) \mid x_1 \in [a_1, b_1], \dots, x_n \in [a_n, b_n]\}.$$

Por exemplo, quando $n = 2$ o produto cartesiano $[\mathbf{a}, \mathbf{b}]$ é um rectângulo.

A demonstração do teorema do valor extremo é paralela à demonstração dada no Volume I para o caso unidimensional. Demonstraremos em primeiro lugar que a continuidade de f implica a *limitação*, para provarmos em seguida que f alcança efectivamente os seus valores máximo e mínimo em $[\mathbf{a}, \mathbf{b}]$.

TEOREMA 9.8. TEOREMA DE LIMITAÇÃO PARA CAMPOS ESCALARES CONTÍNUOS. Se f é um campo escalar contínuo em cada ponto de um intervalo fechado $[\mathbf{a}, \mathbf{b}]$ de \mathbb{R}^n , então f é limitado em $[\mathbf{a}, \mathbf{b}]$, isto é, existe um número $C \geq 0$ tal que $|f(\mathbf{x})| \leq C$ para todo \mathbf{x} em $[\mathbf{a}, \mathbf{b}]$.

Demonstração. Raciocinemos por redução ao absurdo, recorrendo ao método das bissecções sucessivas. A figura 9.11 ilustra o método para o caso em que $n = 2$.

Admitamos que f é não limitado em $[\mathbf{a}, \mathbf{b}]$. Seja $I^{(1)} = [\mathbf{a}, \mathbf{b}]$ e $I_k^{(1)} = [a_k, b_k]$, pelo que

$$I^{(1)} = I_1^{(1)} \times \dots \times I_n^{(1)}.$$

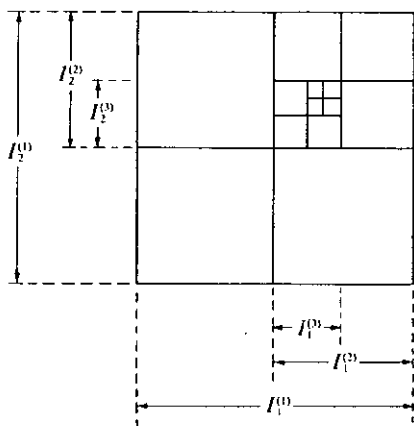


FIG. 9.11. Representação do método da bissecção sucessiva no plano.

Bissectemos cada intervalo unidimensional $I_k^{(1)}$ para formarmos dois subintervalos, a metade esquerda $I_{k,1}^{(1)}$ e a metade direita $I_{k,2}^{(1)}$. Consideremos agora todos os produtos cartesianos possíveis da forma

$$I_{1,j_1}^{(1)} \times \cdots \times I_{n,j_n}^{(1)},$$

onde cada $j_i = 1$ ou 2 . Existem exactamente 2^n produtos deste tipo. Cada um deles é um subintervalo n -dimensional de $[a, b]$, e a sua reunião é igual a $[a, b]$. A função f não é limitada em *um pelo menos* desses subintervalos (se estivesse limitada em todos também o estava em $[a, b]$). Representemos por $I^{(2)}$ um deles que expressamos do modo seguinte

$$I^{(2)} = I_1^{(2)} \times \cdots \times I_n^{(2)},$$

onde cada $I_k^{(2)}$ é um dos subintervalos unidimensionais de $I_k^{(1)}$, de amplitude $\frac{1}{2}(b_k - a_k)$.

Procedamos agora com $I^{(2)}$ como o fizemos com $I^{(1)}$, bissectando cada componente unidimensional $I_k^{(2)}$ e chegando a um intervalo n -dimensional $I^{(3)}$ no qual f é não limitada. Continuamos o processo, obtendo um conjunto infinito de intervalos n -dimensionais

$$I^{(1)}, I^{(2)}, \dots, \quad \text{com} \quad I^{(m+1)} \subseteq I^{(m)},$$

em cada um dos quais f é não limitada. O intervalo $I^{(m)}$ pode representar-se na forma

$$I^{(m)} = I_1^{(m)} \times \cdots \times I_n^{(m)}.$$

Uma vez que cada intervalo unidimensional $I_k^{(m)}$ se obtém por $m - 1$ bissecções sucessivas de $[a_k, b_k]$, fazendo $I_k^{(m)} = [a_k^{(m)}, b_k^{(m)}]$ temos

$$(9.42) \quad b_k^{(m)} - a_k^{(m)} = \frac{b_k - a_k}{2^{m-1}}, \quad \text{para } k = 1, 2, \dots, n.$$

Para cada k fixo, o supremo de todos os extremos esquerdos $a_k^{(m)}$ ($m = 1, 2, \dots$) deve ser portanto igual ao ínfimo de todos os pontos extremos direitos $b_k^{(m)}$ ($m = 1, 2, \dots$) e o seu valor comum representado por t_k . O ponto $t = (t_1, \dots, t_n)$ está em $[a, b]$. Pela continuidade de f em t existe uma n -bola $B(t; r)$ na qual se tem

$$|f(x) - f(t)| \leq 1 \quad \text{para todo } x \text{ em } B(t; r) \cap [a, b].$$

Esta desigualdade implica

$$|f(x)| < 1 + |f(t)| \quad \text{para todo } x \text{ em } B(t; r) \cap [a, b],$$

pelo que f é limitada no conjunto $B(t; r) \cap [a, b]$. Mas este conjunto contém todo o intervalo $I^{(m)}$ quando m é suficientemente grande para que cada um dos n números (9.42) seja menor que r/\sqrt{n} . Por conseguinte para esse valor de m a função f é limitada em $I^{(m)}$, contradizendo o facto de que f é não limitada em $I^{(m)}$. Esta contradição completa a demonstração.

Se f é limitada em $[a, b]$, o conjunto de todos os valores da função $f(x)$ é um conjunto de números reais limitado superior e inferiormente. Portanto este conjunto possui um supremo e um ínfimo os quais se representam por $\sup f$ e $\inf f$, respectivamente. Quer dizer, escrevemos

$$\sup f = \sup \{f(x) \mid x \in [a, b]\}, \quad \inf f = \inf \{f(x) \mid x \in [a, b]\}.$$

Demonstremos agora que uma função contínua torna ambos os valores $\inf f$ e $\sup f$ em $[a, b]$.

TEOREMA 9.9. TEOREMA DOS VALORES EXTREMOS PARA CAMPOS ESCALARES. *Se f é contínua num intervalo fechado $[a, b]$ de \mathbb{R}^n , então existem pontos c e d em $[a, b]$ tais que*

$$f(c) = \sup f \quad \text{e} \quad f(d) = \inf f.$$

Demonstração. Basta provar que f atinge o seu supremo em $[a, b]$. O resultado para o ínfimo aparece então como uma consequência porque o ínfimo de f é o supremo de $-f$.

Seja $M = \sup f$. Suporemos que em $[a, b]$ não existe nenhum x para o qual $f(x) = M$ e chegaremos a um absurdo. Seja $g(x) = M - f(x)$. Para todo x de $[a, b]$ será então $g(x) > 0$, pelo que o recíproco $1/g$ é contínuo em $[a, b]$. Pelo teorema 9.8, $1/g$ é limitada

em $[a, b]$, quer dizer $1/g(x) < C$ para todo x em $[a, b]$, com $C > 0$. Isto implica $M - f(x) > 1/C$, pelo que $f(x) < M - 1/C$ para todo x em $[a, b]$, o que contradiz o facto de M ser o menor limite superior de f em $[a, b]$. Logo $f(x) = M$ para, pelo menos, um x em $[a, b]$.

9.17. O teorema da continuidade uniforme para campos escalares contínuos

Seja f contínua num intervalo limitado fechado $[a, b]$ de \mathbf{R}^n , e designemos por $M(f)$ e $m(f)$, respectivamente, os valores máximo e mínimo de f em $[a, b]$. A diferença

$$M(f) - m(f)$$

chama-se a *oscilação* de f em $[a, b]$. Tal como no caso unidimensional temos um teorema de oscilação uniforme, o qual nos diz que pode efectuar-se uma partição do intervalo $[a, b]$ tal que a oscilação de f em cada subintervalo seja arbitrariamente pequena.

Escrevamos $[a, b] = [a_1, b_1] \times \cdots \times [a_n, b_n]$, e seja P_k uma partição do intervalo $[a_k, b_k]$, isto é, P_k é um conjunto de pontos

$$P_k = \{x_0, x_1, \dots, x_{r-1}, x_r\}$$

tal que $a_k = x_0 \leq x_1 \leq \cdots \leq x_{r-1} \leq x_r = b_k$. O produto cartesiano

$$P = P_1 \times \cdots \times P_n$$

chama-se uma partição do intervalo $[a, b]$. O teorema da oscilação uniforme também chamado o teorema da continuidade uniforme, toma agora a forma seguinte.

TEOREMA 9.10. *Se f é um campo escalar contínuo num intervalo fechado $[a, b]$ de \mathbf{R}^n , então para todo $\epsilon > 0$ existe uma partição de $[a, b]$ num número finito de subintervalos tais que a oscilação de f em cada subintervalo seja menor que ϵ .*

Demonstração. A demonstração é inteiramente análoga ao caso unidimensional pelo que referiremos as passagens principais. Vamos raciocinar por redução ao absurdo, utilizando o método das bissecções sucessivas. Suponhamos que o teorema é falso, isto é, admitamos que para algum ϵ_0 não pode ser feita uma partição do intervalo $[a, b]$ num número finito de subintervalos em cada um dos quais a oscilação de f seja menor que ϵ_0 . Por bissecções sucessivas obtemos um conjunto infinito de subintervalos $I^{(1)}, I^{(2)}, \dots$, em cada um dos quais a oscilação de f é pelo menos ϵ_0 . Considerando o menor limite superior dos extremos esquerdos dos intervalos componentes de $I^{(1)}, I^{(2)}, \dots$ obtemos um ponto t em $[a, b]$ pertencendo a todos estes subintervalos. Pela continuidade de f em t existe uma n -bola $B(t; r)$ tal que a oscilação de f é menor que $\frac{1}{2}\epsilon_0$ em $B(t; r) \cap [a, b]$. Mas, quando m é suficientemente grande, o intervalo $I^{(m)}$ está contido no conjunto $B(t; r) \cap [a, b]$, pelo que a oscilação de f não é maior que $\frac{1}{2}\epsilon_0$ em $I^{(m)}$, o que contradiz o facto de que a oscilação de f é pelo menos ϵ_0 em $I^{(m)}$.

10

INTEGRAIS DE LINHA

10.1. Introdução

No volume I estudámos o integral $\int_a^b f(x)dx$, primeiro para funções reais definidas e limitadas em intervalos finitos, e depois para funções não limitadas e intervalos infinitos. O conceito foi depois generalizado para funções vectoriais e, no Capítulo 7 do Volume II, para funções matriciais.

Neste capítulo generaliza-se a noção de integral noutra direcção. O intervalo $[a, b]$ é substituído por uma curva no espaço n dimensional definida por uma função vectorial α , e o integrando é um campo vectorial f definido e limitado ao longo da curva. O integral chama-se *integral de linha*, *integral curvilíneo*, ou *integral de contorno*, e representa-se por $\int f \cdot d\alpha$, ou por algum outro símbolo semelhante. O ponto usa-se propositadamente para sugerir um produto interno de dois vectores. À curva dá-se o nome de *linha de integração*.

Os integrais de linha são de importância fundamental, quer em Matemática pura, quer aplicada. Aparecem em ligações com as noções de trabalho, energia potencial, fluxo de calor, variação na entropia, circulação de um fluido, e noutras questões físicas em que se estuda o comportamento de um campo escalar ou vectorial ao longo de uma curva.

10.2. Integrais de linha e linhas de integração

Antes de se apresentar a definição de integrais de linha, vamos recordar a definição de curva dada no volume I. Seja α uma função vectorial definida num intervalo finito fechado $J = [a, b]$. Quando t vai tomando os valores J , os valores da função $\alpha(t)$ definem um conjunto de pontos no espaço n dimensional chamados o *gráfico* da função. Se α é contínua em J o gráfico diz-se uma *curva*, mais concretamente a curva descrita por α .

No estudo das curvas efectuado no volume I verificámos que diferentes funções podem dar lugar a uma mesma curva por formas distintas, por exemplo, em diferentes direcções ou com diferentes velocidades. No estudo dos integrais de linha interessamos não somente o conhecimento da curva em si mesma, mas também a maneira como a curva foi originada, isto é, a própria função α . Uma tal função designar-se-á *linha contínua*.

DEFINIÇÃO. Seja $J = [a, b]$ um intervalo finito fechado em \mathbb{R}^1 . Uma função $\alpha: J \rightarrow \mathbb{R}^n$ contínua em J diz-se uma *linha contínua* no espaço n dimensional. A linha diz-se *regular* se a derivada α' existe e é contínua num intervalo aberto (a, b) . A linha diz-se *seccionalmente regular* (regular por intervalos) se pode efectuar-se uma partição do intervalo $[a, b]$ num número finito de subintervalos em cada um dos quais ela seja regular.

A figura 10.1 representa uma curva seccionalmente regular (regular por intervalos). Neste exemplo a curva admite tangente em todos os seus pontos à excepção de um número finito deles. Estes pontos excepcionais subdividem a curva em arcos, ao longo dos quais a tangente à curva muda de direcção de uma maneira contínua.

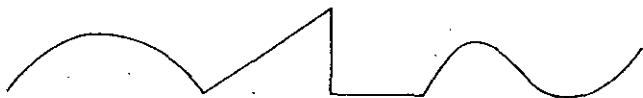


FIG. 10.1. Curva plana seccionalmente regular (regular por intervalos).

DEFINIÇÃO DE INTEGRAL DE LINHA. Seja α uma linha seccionalmente regular no espaço n dimensional definido num intervalo $[a, b]$, e seja f um campo vectorial definido e limitado sobre o gráfico de α . O integral de linha de f ao longo de α representa-se pelo símbolo $\int f \cdot d\alpha$ e define-se por

$$(10.1) \quad \int f \cdot d\alpha = \int_a^b f[\alpha(t)] \cdot \alpha'(t) dt,$$

sempre que o integral do segundo membro existe, quer como integral próprio ou integral impróprio.

Nota: Em muitos dos exemplos que aparecem na prática o produto escalar $f[\alpha(t)] \cdot \alpha'(t)$ é limitado em $[a, b]$ e contínuo, excepto possivelmente num número finito de pontos, caso em que o integral existe como um integral próprio.

10.3. Outras notações para os integrais de linha

Se C representa o gráfico de α , o integral de linha $\int f \cdot d\alpha$ pode também escrever-se $\int_C f \cdot d\alpha$ e chama-se o *integral de f ao longo de C* .

Se $a = \alpha(a)$ e $b = \alpha(b)$ representam as extremidades de C , o integral de linha algumas vezes escreve-se também $\int_a^b f$ ou $\int_a^b f \cdot d\alpha$ e chama-se o *integral de linha de f de*

a a b ao longo de α . Quando se utiliza a notação $\int_a^b f$ deve ter-se em mente que o integral depende não somente dos pontos extremos a e b , mas também da linha α que os une.

Quando $a = b$ a linha diz-se *fechada*. Neste caso usa-se muitas vezes o símbolo \oint para significar a integração ao longo de uma linha fechada.

Quando f e α se exprimem em função das suas componentes, seja

$$f = (f_1, \dots, f_n) \quad \text{e} \quad \alpha = (\alpha_1, \dots, \alpha_n),$$

o integral do segundo membro de (10.1) transforma-se numa soma de integrais,

$$\sum_{k=1}^n \int_a^b f_k[\alpha(t)] \alpha'_k(t) dt.$$

Neste caso o integral de linha pode também escrever-se $\int f_1 d\alpha_1 + \dots + \int f_n d\alpha_n$.

No caso bidimensional a linha α define-se habitualmente por um par de equações paramétricas.

$$x = \alpha_1(t), \quad y = \alpha_2(t),$$

e o integral de linha $\int_C f \cdot d\alpha$ escreve-se $\int_C f_1 dx + f_2 dy$, ou $\int_C f_1(x, y) dx + f_2(x, y) dy$.

No caso tridimensional usam-se as três equações paramétricas,

$$x = \alpha_1(t), \quad y = \alpha_2(t), \quad z = \alpha_3(t)$$

e escrevemos o integral de linha $\int_C f \cdot d\alpha$ na forma $\int_C f_1 dx + f_2 dy + f_3 dz$, ou então

$$\int_C f_1(x, y, z) dx + f_2(x, y, z) dy + f_3(x, y, z) dz.$$

EXEMPLO. Seja f um campo vectorial bidimensional definido por

$$f(x, y) = \sqrt{y} i + (x^3 + y) j$$

para todo (x, y) com $y \geq 0$. Calcular o integral de linha de f de $(0, 0)$ a $(1, 1)$, ao longo de cada um dos seguintes caminhos:

- (a) a recta de equações paramétricas $x = t, y = t, 0 \leq t \leq 1$;
- (b) a curva de equações paramétricas $x = t^2, y = t^3, 0 \leq t \leq 1$.

Resolução. Para a linha da alínea (a) fazemos $\alpha(t) = ti + tj$. Então $\alpha'(t) = i + j$ e $f[\alpha(t)] = \sqrt{t}i + (t^3 + t)j$. Portanto o produto escalar de $f[\alpha(t)]$ e $\alpha'(t)$ vem igual a $\sqrt{t} + t^3 + t$ e encontramos

$$\int_{(0,0)}^{(1,1)} f \cdot d\alpha = \int_0^1 (\sqrt{t} + t^3 + t) dt = \frac{17}{12}.$$

Para o caso da alínea (b) fazemos $\alpha(t) = t^2i + t^3j$. Então $\alpha'(t) = 2ti + 3t^2j$ e $f[\alpha(t)] = t^{3/2}i + (t^6 + t^3)j$. Deste modo

$$f[\alpha(t)] \cdot \alpha'(t) = 2t^{5/2} + 3t^8 + 3t^5,$$

pelo que

$$\int_{(0,0)}^{(1,1)} f \cdot d\alpha = \int_0^1 (2t^{5/2} + 3t^8 + 3t^5) dt = \frac{59}{42}.$$

Este exemplo mostra que o integral de um ponto a outro depende em geral da linha unindo os dois pontos.

Calculemos agora a alínea (b) uma vez mais, usando a mesma curva mas com uma diferente representação paramétrica. A referida curva pode representar-se por

$$\beta(t) = ti + t^{3/2}j, \quad \text{com } 0 \leq t \leq 1.$$

Isto conduz-nos à relação

$$f[\beta(t)] \cdot \beta'(t) = (t^{3/4}i + (t^3 + t^{3/2})j) \cdot (i + \frac{3}{2}t^{1/2}j) = t^{3/4} + \frac{3}{2}t^{7/4} + \frac{3}{2}t^2,$$

cujo integral desde 0 a 1 é $59/42$, como anteriormente. Servem estes cálculos para mostrar que o valor do integral de linha é independente da representação paramétrica usada para definir a curva. É esta uma propriedade geral dos integrais de linha que passamos a demonstrar.

10.4. Propriedades fundamentais dos integrais de linha

Visto os integrais de linha se definirem por intermédio de integrais ordinários, não surpreenderá verificar que gozam de muitas das propriedades destes últimos. Por exemplo, gozam de uma *propriedade de linearidade* a respeito do integrando,

$$\int (af + bg) \cdot d\alpha = a \int f \cdot d\alpha + b \int g \cdot d\alpha,$$

e da *propriedade aditiva* a respeito do caminho de integração:

$$\int_C f \cdot d\alpha = \int_{C_1} f \cdot d\alpha + \int_{C_2} f \cdot d\alpha,$$

onde as duas curvas C_1 e C_2 formam a curva C , quer dizer, a curva C é definida por uma função α definida no intervalo $[a, b]$, e as curvas C_1 e C_2 são as definidas por $\alpha(t)$ quando t varia nos subintervalos $[a, c]$ e $[c, b]$, respectivamente, para algum c satisfazendo a $a < c < b$. As demonstrações destas propriedades resultam directamente da definição de integral de linha e como tal são deixados ao leitor como exercício.

Seguidamente examinaremos o comportamento dos integrais de linha ao efectuar-se uma mudança de parâmetro. Sejam α uma linha contínua definida num intervalo $[a, b]$, a u uma função real derivável, com u' nunca nula num intervalo $[c, d]$, e tal que o contradomínio de u seja $[a, b]$. Então a função β definida em $[c, d]$ pela equação

$$\beta(t) = \alpha[u(t)]$$

é uma linha contínua que tem o mesmo gráfico que α . Duas linhas α e β assim relacionados dizem-se *equivalentes*. Diz-se que proporcionam representações paramétricas distintas da mesma curva. A função u define uma mudança de parâmetro.

Seja C o gráfico comum das duas linhas equivalentes α e β . Se a derivada de u for sempre positiva em $[c, d]$, a função u é crescente e dizemos então que as duas linhas α e β correspondem a C com a *mesmo sentido*. Se a derivada de u for sempre negativa, dizemos que a α e β correspondem a C com *sentidos opostos*. No primeiro caso diz-se que a função u *preserva a orientação*; no segundo caso diz-se que u *inverte a orientação*. Na figura 10.2 apresentam-se exemplos.

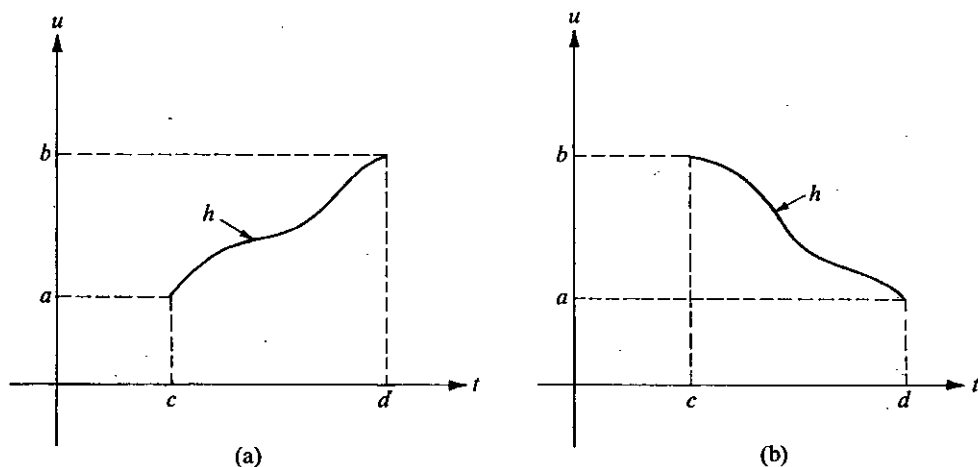


FIG. 10.2. Uma mudança de parâmetro definida por $u = h(t)$. Em (a), a função h preserva a orientação. Em (b), a função h inverte a orientação.

O teorema que estudamos a seguir mostra que um integral de linha não se altera sob uma mudança do parâmetro que preserve a orientação; muda de sinal se a mudança de parâmetro inverte a orientação. Supomos que ambos os integrais $\int f \cdot d\alpha$ e $\int f \cdot d\beta$ existem.

TEOREMA 10.1. COMPORTAMENTO DE UM INTEGRAL DE LINHA SOB UMA MUDANÇA DE PARÂMETRO. Se α e β são duas linhas seccionalmente regulares, então tem-se

$$\int_C f \cdot d\alpha = \int_C f \cdot d\beta$$

se α e β originam C com o mesmo sentido, e

$$\int_C f \cdot d\alpha = - \int_C f \cdot d\beta$$

se α e β originam C com sentidos opostos.

Demonstração. Basta demonstrar o teorema para linhas regulares, uma vez que podemos depois invocar a propriedade aditiva a respeito da linha de integração para deduzir o resultado para linhas seccionalmente regulares.

A demonstração é uma aplicação simples da generalização da regra de derivação da função composta. As linhas α e β estão relacionadas por uma equação da forma $\beta(t) = \alpha[u(t)]$, onde u se define num intervalo $[c, d]$ e α se define em $[a, b]$. Pela regra referida temos

$$\beta'(t) = \alpha'[u(t)]u'(t).$$

Portanto temos

$$\int_C f \cdot d\beta = \int_c^d f[\beta(t)] \cdot \beta'(t) dt = \int_c^d f[\alpha(u(t))] \cdot \alpha'[u(t)]u'(t) dt.$$

No último integral introduzimos a substituição $v = u(t)$, $dv = u'(t)dt$ a fim de obtermos

$$\int_C f \cdot d\beta = \int_{u(c)}^{u(d)} f(\alpha(v)) \cdot \alpha'(v) dv = \pm \int_a^b f(\alpha(v)) \cdot \alpha'(v) dv = \pm \int_C f \cdot d\alpha,$$

onde se utiliza o sinal $+$ se $a = u(c)$ e $b = u(d)$, e o sinal $-$ se $a = u(d)$ e $b = u(c)$. O primeiro caso verifica-se se α e β originam C na mesmo sentido lo segundo se originam C em sentidos opostos.

10.5. Exercícios

Em cada um dos Exercícios 1 a 8 calcular o integral de linha do campo vectorial f ao longo da linha indicada.

- $f(x, y) = (x^2 - 2xy)i + (y^2 - 2yx)j$, de $(-1, 1)$ a $(1, 1)$ ao longo da parábola $y = x^2$.
- $f(x, y) = (2a - y)i + xj$, ao longo da linha definida por $\alpha(t) = a(t - \sin t)i + a(1 - \cos t)j$, $0 \leq t \leq 2\pi$.
- $f(x, y, z) = (y^2 - z^2)i + 2yzj - x^2k$, ao longo da linha definida por $\alpha(t) = ti + t^2j + t^3k$, $0 \leq t \leq 1$.
- $f(x, y) = (x^2 + y^2)i + (x^2 - y^2)j$, de $(0, 0)$ a $(2, 0)$ ao longo da curva $y = 1 - (1 - x)$.
- $f(x, y) = (x + y)i + (x - y)j$, ao longo da elipse completa $b^2x^2 + a^2y^2 = a^2b^2$, no sentido directo.
- $f(x, y, z) = 2xyi + (x^2 + z)j + yk$, de $(1, 0, 2)$ a $(3, 4, 1)$ ao longo do segmento de recta que definem.
- $f(x, y, z) = xi + yj + (xz - y)k$, de $(0, 0, 0)$ a $(1, 2, 4)$. Ao longo do segmento de recta que definem.

8. $f(x, y, z) = xi + yj + (xz - y)k$, ao longo da linha definida por $\alpha(t) = t^2i + 2tj + 4t^3k$, $0 \leq t \leq 1$.
 Em cada um dos Exercícios 9 a 12, calcular o valor do integral de linha dado.
9. $\int_C (x^2 - 2xy)dx + (y^2 - 2xy)dy$, onde C é a parábola $y = x^2$ de $(-2, 4)$ a $(1, 1)$.
10. $\int_C \frac{(x+y)dx - (x-y)dy}{x^2 + y^2}$, onde C é o círculo $x^2 + y^2 = a^2$, percorrido uma vez no sentido directo.
11. $\int_C \frac{dx + dy}{|x| + |y|}$, onde C é o quadrado de vértices $(1, 0)$, $(0, 1)$, $(-1, 0)$, e $(0, -1)$, percorrido uma vez no sentido directo.
12. $\int_C y dx + z dy + x dz$, onde
 (a) C é a curva de intersecção de duas superfícies $x + y = 2$ e $x^2 + y^2 + z^2 = 2(x + y)$. A curva é percorrida uma vez no sentido retrógrado quando vista da origem.
 (b) C é a intersecção de duas superfícies $z = xy$ e $x^2 + y^2 = 1$, percorrida uma vez no sentido retrógrado observado acima de XOY .

10.6. O conceito de trabalho como um integral de linha

Consideremos uma partícula movendo-se ao longo de uma curva sob a acção de um campo de forças f . Se a curva é a representação gráfica de α , seccionalmente regular, o trabalho produzido por f define-se através do integral de linha $\int f \cdot d\alpha$. Os exemplos que se seguem ilustram algumas propriedades fundamentais do trabalho.

EXEMPLO 1. *Trabalho produzido por uma força constante.* Se f for uma força constante, por exemplo $f = c$, pode provar-se que o trabalho realizado por f ao deslocar uma partícula de um ponto a para um ponto b , ao longo de qualquer trajectória seccionalmente regular unindo a e b é $c \cdot (b - a)$, o produto escalar da força pelo vector deslocamento $b - a$. Vamos demonstrar esta afirmação num caso particular.

Seja $\alpha = (\alpha_1, \dots, \alpha_n)$ uma linha unindo a e b , a saber $\alpha(a) = a$ e $\alpha(b) = b$, e seja $c = (c_1, \dots, c_n)$. Admitindo que α' é contínua em $[a, b]$, então o trabalho realizado por f é igual a

$$\int f \cdot d\alpha = \sum_{k=1}^n c_k \int_a^b \alpha'_k(t) dt = \sum_{k=1}^n c_k [\alpha_k(b) - \alpha_k(a)] = c \cdot [\alpha(b) - \alpha(a)] = c \cdot (b - a).$$

Para este campo de forças o trabalho depende unicamente dos pontos extremos a e b e não da curva que os une. Porém nem todos os campos de forças gozam desta propriedade. Aqueles para os quais ela se verifica dizem-se *conservativos*. O exemplo da página 365 é o de um campo de forças não conservativo. Numa secção posterior determinaremos todos os campos de forças conservativos.

EXEMPLO 2. *O princípio do trabalho e energia.* Uma partícula de massa m move-se ao longo de uma curva sob a acção de um campo de forças f . Se a velocidade da partícula no instante t é $v(t)$, a sua energia cinética é $\frac{1}{2}mv^2(t)$. Provar que a *variação da ener-*

gia cinética em determinado intervalo de tempo é igual ao trabalho realizado por f nesse intervalo de tempo.

Resolução. Seja $\mathbf{r}(t)$ a posição da partícula no instante t . O trabalho efectuado por f durante o intervalo de tempo $[a, b]$ é $\int_a^b f \cdot d\mathbf{r}$. Pretendemos provar que

$$\int_{\mathbf{r}(a)}^{\mathbf{r}(b)} f \cdot d\mathbf{r} = \frac{1}{2}mv^2(b) - \frac{1}{2}mv^2(a).$$

Pela segunda lei de Newton do movimento temos

$$f[\mathbf{r}(t)] = m\mathbf{r}''(t) = m\mathbf{v}'(t),$$

representando $\mathbf{v}(t)$ o vector velocidade no instante t . A grandeza do vector velocidade é $v(t) = \|\mathbf{v}(t)\|$. Deste modo

$$f[\mathbf{r}(t)] \cdot \mathbf{r}'(t) = f[\mathbf{r}(t)] \cdot \mathbf{v}(t) = m\mathbf{v}'(t) \cdot \mathbf{v}(t) = \frac{1}{2}m \frac{d}{dt} (\mathbf{v}(t) \cdot \mathbf{v}(t)) = \frac{1}{2}m \frac{d}{dt} (v^2(t)).$$

Integrando de a e b obtemos

$$\int_{\mathbf{r}(a)}^{\mathbf{r}(b)} f \cdot d\mathbf{r} = \int_a^b f[\mathbf{r}(t)] \cdot \mathbf{r}'(t) dt = \left. \frac{1}{2}mv^2(t) \right|_a^b = \frac{1}{2}mv^2(b) - \frac{1}{2}mv^2(a),$$

como se pretendia demonstrar.

10.7. Integrais de linha relativos ao comprimento de arco

Seja α uma linha com derivada α' continua num intervalo $[a, b]$. O gráfico de α é uma curva rectificável. No Volume I provámos que a correspondente função comprimento do arco s é definida pelo integral

$$s(t) = \int_a^t \|\alpha'(u)\| du.$$

A derivada do comprimento de arco é dada por

$$s'(t) = \|\alpha'(t)\|.$$

Seja φ um campo escalar definido e limitado em C , o gráfico de α . O integral de linha de φ a respeito do comprimento do arco ao longo de C representa-se pelo símbolo $\int_C \varphi ds$ e define-se

$$\int_C \varphi ds = \int_a^b \varphi[\alpha(t)]s'(t) dt,$$

sempre que o integral do segundo membro existe.

Consideremos agora um campo escalar φ definido por $\varphi[\alpha(t)] = f[\alpha(t)]$. $T(t)$, o producto escalar de um campo vectorial f definido sobre C pelo vector unitário tangen-

te $T(t) = (d\alpha/ds)$. Nesta hipótese o integral de linha $\int_C \varphi ds$ é o mesmo que o integral $\int_C f \cdot d\alpha$ porque

$$f[\alpha(t)] \cdot \alpha'(t) = f[\alpha(t)] \cdot \frac{d\alpha}{ds} \frac{ds}{dt} = f[\alpha(t)] \cdot T(t) s'(t) = \varphi[\alpha(t)] s'(t).$$

Quando f representa o campo das velocidades, o produto escalar $f \cdot T$ dá a componente tangencial da velocidade, e o integral de linha $\int_C f \cdot T ds$ é o chamado *integral do fluxo* de f ao longo de C . Quando C é uma curva fechada o integral do fluxo diz-se a *circulação* de f ao longo de C . Estes termos são usuais na teoria do fluxo de fluidos.

10.8. Outras aplicações dos integrais de linha

Os integrais de linha com respeito ao comprimento de arco ocorrem também em problemas relacionados com a distribuição de massas ao longo de uma curva. Por exemplo, imaginemos uma curva C no espaço tridimensional materializada por um fio de arame de densidade variável. Admitamos que a densidade é definida por um campo escalar φ , sendo $\varphi(x, y, z)$ a *massa por unidade de comprimento* no ponto (x, y, z) de C . Então a massa total M do fio é dada pelo integral de linha de φ relativo ao comprimento de arco:

$$M = \int_C \varphi(x, y, z) ds.$$

O *centro de massa* do fio é o ponto $(\bar{x}, \bar{y}, \bar{z})$ cujas coordenadas são dadas pelas fórmulas

$$\bar{x}M = \int_C x\varphi(x, y, z) ds, \quad \bar{y}M = \int_C y\varphi(x, y, z) ds, \quad \bar{z}M = \int_C z\varphi(x, y, z) ds.$$

Um fio de densidade constante diz-se *homogêneo*. Neste caso o centro de massa chama-se também o *centroide*.

EXEMPLO 1. Determinar a massa M de uma espira completa de uma mola em forma de hélice cilíndrica de equação vectorial

$$\alpha(t) = a \cos t \mathbf{i} + a \sin t \mathbf{j} + btk$$

se a densidade em (x, y, z) for $x^2 + y^2 + z^2$.

Resolução. O integral que define M é

$$M = \int_C (x^2 + y^2 + z^2) ds = \int_0^{2\pi} (a^2 \cos^2 t + a^2 \sin^2 t + b^2 t^2) s'(t) dt.$$

Porque $s'(t) = \|\alpha'(t)\|$ e $\alpha'(t) = -a \sin t \mathbf{i} + a \cos t \mathbf{j} + b \mathbf{k}$, tem-se $s'(t) = \sqrt{a^2 + b^2}$ e consequente

$$M = \sqrt{a^2 + b^2} \int_0^{2\pi} (a^2 + b^2 t^2) dt = \sqrt{a^2 + b^2} \left(2\pi a^2 + \frac{8}{3} \pi^3 b^2 \right).$$

Neste exemplo a coordenada \bar{z} do centro de massa é dada por

$$\begin{aligned} \bar{z}M &= \int_C z(x^2 + y^2 + z^2) ds = \sqrt{a^2 + b^2} \int_0^{2\pi} bt(a^2 + b^2 t^2) dt \\ &= b\sqrt{a^2 + b^2} (2\pi^2 a^2 + 4\pi^4 b^2). \end{aligned}$$

A determinação das coordenadas \bar{x} e \bar{y} é pedida no Exercício 15 da Secção 10.9.

Os integrais de linha podem utilizar-se para definir o momento de inércia de um fio em relação a um eixo. Se $\delta(x, y, z)$ representa a distância de um ponto (x, y, z) de C um eixo L , o momento de inércia I_L define-se pelo integral

$$I_L = \int_C \delta^2(x, y, z) \varphi(x, y, z) ds,$$

em que $\varphi(x, y, z)$ é a densidade em (x, y, z) . Os momentos de inércia em relação aos eixos coordenados representam-se por I_x , I_y e I_z .

EXEMPLO 2. Calcular o momento de inércia I_z da espira completa de mola helicoidal referida no Exemplo 1.

Resolução. Aqui $\delta^2(x, y, z) = x^2 + y^2 = a^2$ e $\varphi(x, y, z) = x^2 + y^2 + z^2$, pelo que se tem

$$I_z = \int_C (x^2 + y^2)(x^2 + y^2 + z^2) ds = a^2 \int_C (x^2 + y^2 + z^2) ds = Ma^2,$$

onde M é a massa que se calculou no Exemplo 1.

10.9. Exercícios

1. Um campo de forças f , no espaço tridimensional, é definido por $f(x, y, z) = x\mathbf{i} + y\mathbf{j} + (xz - y)\mathbf{k}$. Calcular o trabalho efectuado por esta força ao deslocar uma partícula de $(0, 0, 0)$ para $(1, 2, 4)$, ao longo do segmento de recta unindo estes dois pontos.
2. Determinar o trabalho total efectuado pela força $f(x, y) = (x^2 - y^2)\mathbf{i} + 2xy\mathbf{j}$ ao deslocar uma partícula (no sentido directo) percorrendo uma vez o quadrado definido pelos eixos coordenados e pelas rectas $x = a$ e $y = a$, $a > 0$.
3. Um campo de forças bidimensional f é definido pela equação $f(x, y) = cxy\mathbf{i} + x^4y^2\mathbf{j}$, com c uma constante positiva. Esta força actua sobre uma partícula fazendo deslocá-la desde $(0, 0)$ até à recta $x = 1$ ao longo de uma curva definida por

$$y = ax^b, \quad \text{onde } a > 0 \quad \text{e} \quad b > 0.$$

Determinar um valor de a (em função de c) tal que o trabalho efectuado por esta força seja independente de b .

4. Um campo de forças f , no espaço tridimensional, é dado pela fórmula $f(x, y, z) = yzi + xzj + x(y + 1)k$. Calcular o trabalho efectuado por f ao deslocar uma partícula percorrendo uma vez o contorno do triângulo de vértices $(0, 0, 0)$, $(1, 1, 1)$, $(-1, 1, -1)$ por esta ordem.
5. Calcular o trabalho efectuado por um campo de forças $f(x, y, z) = (y - z)i + (z - x)j + (x - y)k$ ao longo da curva de intersecção da esfera $x^2 + y^2 + z^2 = 4$ e o plano $z = y \tan \theta$, com $0 < \theta < \pi/2$. A trajectória é percorrida no sentido directo para um observador perpendicular a XOY segundo OZ positivo.
6. Calcular o trabalho efectuado pelo campo de forças $f(x, y, z) = y^2i + z^2j + x^2k$ ao longo da curva de intersecção da esfera $x^2 + y^2 + z^2 = a^2$ e o cilindro $x^2 + y^2 = ax$, com $z \geq 0$ e $a > 0$. A trajectória é percorrida no sentido retrógrado quando observada do semi-eixo positivo OZ .
Calcular os integrais de linha, com respeito ao comprimento de arco, nos Exercícios 7 a 10.
7. $\int_C (x + y)ds$, onde C é o triângulo de vértices $(0, 0)$, $(1, 0)$ e $(0, 1)$, percorrida no sentido directo.
8. $\int_C y^2 ds$, onde C tem a equação vectorial

$$\alpha(t) = a(t - \sin t)i + a(1 - \cos t)j, \quad 0 \leq t \leq 2\pi.$$

9. $\int_C (x^2 + y^2)ds$, onde C tem a equação vectorial

$$\alpha(t) = a(\cos t + t \sin t)i + a(\sin t - t \cos t)j, \quad 0 \leq t \leq 2\pi.$$

10. $\int_C z ds$, onde C tem a equação vectorial

$$\alpha(t) = t \cos t i + t \sin t j + tk, \quad 0 \leq t \leq t_0.$$

11. Considerar um fio homogéneo de forma semicircular de raio a .
(a) Provar que o centro de massa está situado sobre o eixo de simetria a uma distância de $2a/\pi$ do centro.
(b) Provar que o momento de inércia em relação ao diâmetro passando pelas extremidades do fio é $\frac{1}{2}Ma^2$, onde M é a massa do fio.
12. Um fio tem a forma do círculo $x^2 + y^2 = a^2$. Determinar a sua massa e o momento de inércia em relação a um diâmetro, sabendo que a densidade em (x, y) é $|x| + |y|$.
13. Determinar a massa de um fio cuja forma é a da curva de intersecção da esfera $x^2 + y^2 + z^2 = 1$ pelo plano $x + y + z = 0$, se a densidade do fio em (x, y, z) é x^2 .
14. Um fio homogéneo tem a forma da porção da curva de intersecção da superfície $x^2 + y^2 = z^2$ com a superfície $y^2 = x$, compreendida entre os pontos $(0, 0, 0)$ e $(1, 1, \sqrt{2})$. Determinar a cota \bar{z} do seu centro de massa.
15. Determinar as coordenadas \bar{x} e \bar{y} do centro de massa de uma espira completa da mola referida no Exemplo 1 da secção 10.8.
16. Para a espira completa da mola referida no Exemplo 1 da Secção 10.8, calcular os momentos de inércia I_x e I_y .

10.10. Conjuntos conexos abertos. Independência da linha

Seja S um conjunto aberto de \mathbb{R}^n . O conjunto S diz-se *conexo* (por arcos) se cada par de pontos de S pode unir-se mediante uma linha seccionalmente regular cujo gráfico esteja situado em S , isto é, para cada par de pontos a e b em S exista uma linha seccionalmente regular α , definida num intervalo $[a, b]$, tal que $\alpha(t) \in S$ para cada t de $[a, b]$, com $\alpha(a) = a$ e $\alpha(b) = b$.

Na figura 10.3 estão representados três exemplos de conjuntos abertos conexos no plano. Exemplos correspondentes a estes, no espaço tridimensional seriam (a) um elipsoide, (b) um poliedro, e (c) um toro; em cada caso só são considerados os pontos interiores.

Um conjunto aberto S diz-se que é *não conexo* se S é a união de dois ou mais con-

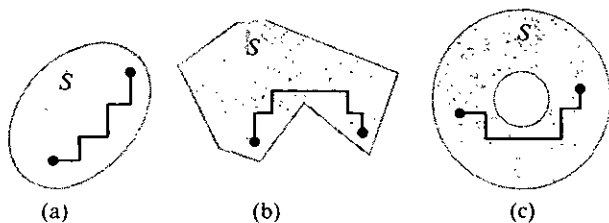


FIG. 10.3. Exemplos de conjuntos abertos conexos

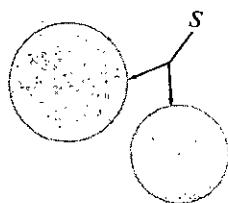


FIG. 10.4. Um conjunto não conexo S , a união de dois discos circulares disjuntos.

conjuntos abertos não vazios e disjuntos, de que se apresenta um exemplo na figura 10.4. Pode mostrar-se que a classe de conjuntos abertos conexos é idêntica à dos conjuntos abertos não conexos†.

Seja f um campo vectorial contínuo num conjunto aberto conexo S . Sejam a e b dois pontos de S e consideremos o integral de linha de f de a e b ao longo de certa linha de S , seccionalmente regular. O valor do integral depende, em geral, da linha unindo a a b . Para certos campos vectoriais, o integral depende unicamente dos pontos extremos a e b e não da linha que os une. Neste caso dizemos que o *integral é independente da linha unindo a a b* . Dizemos que o integral de linha de f é *independente da linha ao longo da qual se define em S* se é independente da linha unindo a a b para todo o par de pontos de S .

Para que campos vectoriais são os integrais de linha independentes da linha ao longo da qual se definem? Para responder a esta questão vamos generalizar o primeiro e segundo teoremas fundamentais do cálculo aos integrais de linha.

10.11. O segundo teorema fundamental do cálculo para os integrais de linha

O segundo teorema fundamental para funções reais, como se demonstrou no Volume I (Teorema 5.3), estabelece que

† Para um estudo mais profundo da conexão de conjuntos, ver o cap. 8 da obra do mesmo autor *Mathematical Analysis*, Addison-Wesley (edição espanhola *Análisis matemático*, Editorial Reverté, S.A., Barcelona).

$$\int_a^b \varphi'(t) dt = \varphi(b) - \varphi(a),$$

desde que φ' seja contínua em algum intervalo aberto contendo a e b . Para generalizar este resultado aos integrais de linha necessitamos de uma versão ligeiramente mais forte do teorema em que a continuidade de φ' se supõe unicamente no intervalo aberto (a, b) .

TEOREMA 10.2. *Se φ é uma função real contínua num intervalo fechado $[a, b]$ e se se admite que o integral $\int_a^b \varphi'(t) dt$ existe e φ' é contínuo no intervalo aberto (a, b) , tem-se*

$$\int_a^b \varphi'(t) dt = \varphi(b) - \varphi(a).$$

Demonstração. Para cada x em $[a, b]$ definimos $f(x) = \int_a^x \varphi'(t) dt$. Pretendemos provar que

$$(10.2) \quad f(b) = \varphi(b) - \varphi(a).$$

Pelo teorema 3.4 do Volume I, f é contínua num intervalo fechado $[a, b]$. Pelo teorema 5.1 do Volume I, f é derivável no intervalo aberto (a, b) com $f'(x) = \varphi'(x)$ para todo x em (a, b) . Deste modo, pelo teorema 5.2 do Volume I, a diferença $f - \varphi$ é constante no intervalo aberto (a, b) . Devido à continuidade, $f - \varphi$ é também constante no intervalo fechado $[a, b]$. Em particular, $f(b) - \varphi(b) = f(a) - \varphi(a)$. Mas como $f(a) = 0$, fica demonstrada (10.2).

TEOREMA 10.3. SEGUNDO TEOREMA FUNDAMENTAL DO CÁLCULO PARA INTEGRAIS DE LINHA: *Se φ é um campo escalar diferenciável, com gradiente contínuo $\nabla \varphi$ num conjunto conexo aberto S de \mathbb{R}^n , então para dois quaisquer pontos a e b ligados por uma linha seccionalmente regular α em S tem-se*

$$\int_a^b \nabla \varphi \cdot d\alpha = \varphi(b) - \varphi(a).$$

Demonstração. Escolhamos dois pontos quaisquer a e b em S e unamo-los por uma linha seccionalmente regular α em S definida num intervalo $[a, b]$. Suponhamos em primeiro lugar que α é regular em $[a, b]$. Então o integral de linha de $\nabla \varphi$ de a a b ao longo de α é dado por

$$\int_a^b \nabla \varphi \cdot d\alpha = \int_a^b \nabla \varphi[\alpha(t)] \cdot \alpha'(t) dt.$$

Mas

$$\nabla \varphi[\alpha(t)] \cdot \alpha'(t) = g'(t),$$

onde g é a função composta definida em $[a, b]$ pela fórmula

$$g(t) = \varphi[\alpha(t)].$$

A derivada g' é contínua no intervalo aberto (a, b) porque $\nabla\varphi$ é contínuo em S e α é regular. Portanto podemos aplicar o teorema 10.2 a g obtendo

$$\int_a^b \nabla\varphi \cdot d\alpha = \int_a^b g'(t) dt = g(b) - g(a) = \varphi[\alpha(b)] - \varphi[\alpha(a)] = \varphi(b) - \varphi(a).$$

Isto demonstra o teorema se α é regular.

Quando α é seccionalmente regular efectuamos uma partição do intervalo $[a, b]$ num número finito (por exemplo r) de subintervalos $[t_{k-1}, t_k]$, em cada um dos quais α seja regular e aplicamos o que acabámos de demonstrar a cada um dos subintervalos. Obtemos então

$$\int_a^b \nabla\varphi = \sum_{k=1}^r \int_{\alpha(t_{k-1})}^{\alpha(t_k)} \nabla\varphi = \sum_{k=1}^r \{\varphi[\alpha(t_k)] - \varphi[\alpha(t_{k-1})]\} = \varphi(b) - \varphi(a),$$

Como pretendíamos demonstrar.

Como uma consequência do teorema 10.3 vemos que o integral de linha de um gradiente é independente da linha em qualquer conjunto aberto conexo S no qual o gradiente seja contínuo. Para uma linha fechada temos $b = a$; pelo que $\varphi(b) - \varphi(a) = 0$. Por outras palavras, *o integral de linha de um gradiente contínuo é zero ao longo de toda a linha fechada seccionalmente regular situada em S* . Na Secção 10.14 provaremos (teorema 10.4) que os gradientes são os *únicos* campos vectoriais contínuos gozando desta propriedade.

10.12. Aplicações à mecânica

Se um campo vectorial f é o gradiente de um campo escalar φ , então φ chama-se uma *função potencial* para f . No espaço tridimensional, os conjuntos de nível de φ chamam-se *superfícies equipotenciais*; no espaço bidimensional dizem-se *linhas equipotenciais*. (Se φ representa a temperatura, a palavra “equipotencial” aparece substituída por “isotérmica”; se φ representa a pressão aparece a palavra “isobárica”).

EXEMPLO 1. No espaço tridimensional, seja $\varphi(x, y, z) = r^n$, com $r = (x^2 + y^2 + z^2)^{1/2}$. Para cada inteiro n temos

$$\nabla(r^n) = nr^{n-2}\mathbf{r},$$

onde $\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$. (Ver Exercício 18 da Secção 8.14.) Por conseguinte φ é um potencial do campo vectorial

$$f(x, y, z) = nr^{n-2}\mathbf{r}.$$

As superfícies equipotenciais de φ são esferas centradas na origem.

EXEMPLO 2. *O potencial newtoniano.* A lei de gravitação de Newton estabelece que a força f que numa partícula de massa M exerce sobre outra partícula de massa m é um vector de grandeza GmM/r^2 , sendo G uma constante e r a distância entre as duas partículas. Definindo a origem na partícula de massa M e sendo $r = xi + yj + zk$ o vector posicional, em relação àquela origem, da partícula de massa m , então $r = \|r\|$ e $-r/r$ é um vector unitário com o mesmo sentido que f , pelo que a lei de Newton toma a forma

$$f = -GmMr^{-3}r.$$

Fazendo $n = -1$ no Exemplo 1, vemos que a força gravitacional f é o gradiente do campo escalar dado por

$$\varphi(x, y, z) = GmMr^{-1}.$$

Este é o *potencial newtoniano*.

O trabalho efectuado pela força gravitacional ao deslocar uma partícula de massa m de (x_1, y_1, z_1) a (x_2, y_2, z_2) é

$$\varphi(x_1, y_1, z_1) - \varphi(x_2, y_2, z_2) = GmM \left(\frac{1}{r_1} - \frac{1}{r_2} \right),$$

onde $r_1 = (x_1^2 + y_1^2 + z_1^2)^{1/2}$ e $r_2 = (x_2^2 + y_2^2 + z_2^2)^{1/2}$. Se os dois pontos pertencem à mesma superfície equipotencial então $r_1 = r_2$ e o trabalho é nulo.

EXEMPLO 3. *O princípio da conservação da energia mecânica.* Seja f um campo de forças contínuas admitindo um potencial φ num conjunto aberto conexo S . O teorema 10.3 diz-nos que o trabalho efectuado por f ao deslocar uma partícula de a para x , ao longo de qualquer linha seccionalmente regular em S , é $\varphi(x) - \varphi(a)$, a variação da função potencial. No Exemplo 2 da Secção 10.6 provámos que este trabalho é igual a variação da energia cinética da partícula, $k(x) - k(a)$, representando $k(x)$ a energia cinética da partícula quando colocada em x . Deste modo, temos

$$k(x) - k(a) = \varphi(x) - \varphi(a),$$

ou

$$(10.3) \quad k(x) - \varphi(x) = k(a) - \varphi(a).$$

O escalar $- \varphi(x)$ chama-se a *energia potencial*† da partícula.

Se a se considera fixo e se faz variar x no conjunto S , a equação (10.3) diz-nos que a soma de $k(x)$ com $- \varphi(x)$ é constante. Quer dizer, *se uma campo de forças f é um gra-*

† Alguns autores referem-se a $- \varphi$ como sendo a função potencial de f pelo que a energia potencial em x será igual ao valor da função potencial φ em x .

diente, a soma das energias cinética e potencial de uma partícula movendo-se neste campo é constante. Em Mecânica este é o chamado princípio de conservação da energia (mecânica). Um campo de forças admitindo uma função potencial diz-se *conservativo* porque a energia total, soma das energias cinética e potencial, se conserva. Num campo conservativo não se realiza trabalho algum no movimento de uma partícula ao longo de uma curva fechada voltando ao ponto de partida. Um campo de forças não será conservativo se no sistema existirem atrito ou viscosidade, uma vez que estes tendem a transformar energia mecânica em energia calorífica.

10.13. Exercícios

- Determinar quais dos seguintes conjuntos abertos S de \mathbb{R}^2 são conexos. Para cada conjunto conexo definir dois pontos arbitrários distintos de S e explicar como se poderia encontrar em S uma curva seccionalmente regular de S ligando os dois pontos.
 - $S = \{(x, y) \mid x^2 + y^2 \geq 0\}$.
 - $S = \{(x, y) \mid x^2 + y^2 > 0\}$.
 - $S = \{(x, y) \mid x^2 + y^2 < 1\}$.
 - $S = \{(x, y) \mid 1 < x^2 + y^2 < 2\}$.
 - $S = \{(x, y) \mid x^2 + y^2 > 1 \text{ e } (x-3)^2 + y^2 > 1\}$.
 - $S = \{(x, y) \mid x^2 + y^2 < 1 \text{ ou } (x-3)^2 + y^2 < 1\}$.
- Dado um campo vectorial bidimensional

$$f(x, y) = P(x, y)\mathbf{i} + Q(x, y)\mathbf{j},$$

no qual as derivadas parciais $\partial P/\partial y$ e $\partial Q/\partial x$ são contínuas num conjunto aberto S , se f é o gradiente de certo potencial ϕ , provar que

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}$$

em todo o ponto de S .

- Para cada um dos campos vectoriais seguintes, utilizar o resultado do Exercício 2 para provar que f não é um gradiente. Determinar então uma curva fechada C tal que $\int_C f \neq 0$.

(a) $f(x, y) = y\mathbf{i} - x\mathbf{j}$.

(b) $f(x, y) = y\mathbf{i} + (xy - x)\mathbf{j}$.

- Dado um campo vectorial tridimensional

$$f(x, y, z) = P(x, y, z)\mathbf{i} + Q(x, y, z)\mathbf{j} + R(x, y, z)\mathbf{k},$$

em que as derivadas parciais

$$\frac{\partial P}{\partial y}, \frac{\partial P}{\partial z}, \frac{\partial Q}{\partial x}, \frac{\partial Q}{\partial z}, \frac{\partial R}{\partial x}, \frac{\partial R}{\partial y},$$

são contínuas num conjunto aberto S , se f é o gradiente de alguma função potencial ϕ , provar que

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}, \quad \frac{\partial P}{\partial z} = \frac{\partial R}{\partial x}, \quad \frac{\partial Q}{\partial z} = \frac{\partial R}{\partial y}$$

em cada ponto de S .

5. Para cada um dos seguintes campos vectoriais, aplicar o resultado do Exercício 4 para provar que f não é um gradiente. Determinar depois numa curva fechada C tal que $\oint_C f \neq 0$.

(a) $f(x, y, z) = yi + xj + xk$.

(b) $f(x, y, z) = xyi + (x^2 + 1)j + z^2k$.

6. No espaço tridimensional um campo de forças f define-se pela equação

$$f(x, y, z) = yi + zj + yzk.$$

(a) Determinar se f é ou não conservativo

(b) Calcular o trabalho realizado ao deslocar uma partícula ao longo da curva definida por

$$\alpha(t) = \cos t i + \sin t j + e^t k$$

quando t varia de 0 a π .

7. Um campo de forças bidimensional F define-se por

$$F(x, y) = (x + y)i + (x - y)j.$$

(a) Provar que o trabalho realizado por esta força ao deslocar uma partícula ao longo da curva

$$\alpha(t) = f(t)i + g(t)j, \quad a \leq t \leq b,$$

depende unicamente de $f(a)$, $f(b)$, $g(a)$ e $g(b)$.

(b) Determinar o trabalho total realizado quando $f(a) = 1$, $f(b) = 2$, $g(a) = 3$, $g(b) = 4$.

8. Um campo de forças define-se em coordenadas polares pela equação

$$F(r, \theta) = -4 \sin \theta i + 4 \sin \theta j.$$

Calcular o trabalho realizado no deslocamento de uma partícula do ponto $(1, 0)$ até à origem, ao longo da espiral de equação $r = e^{-\theta}$.

9. Um campo de forças "central" F no plano pode definir-se por $F(x, y) = f(r)r$, com $r = xi + yj$ e $r = \|r\|$. Provar que tal campo de forças é conservativo.

10. Determinar o trabalho realizado pela força $F(x, y) = (3y^2 + 2)i + 16xj$ no deslocamento de uma partícula de $(-1, 0)$ a $(1, 0)$ ao longo da semi-ellipse superior de equação $b^2x^2 + y^2 = b^2$. Para que ellipse (isto é, para que valor de b) será mínimo o trabalho realizado?

10.14. O primeiro teorema fundamental do cálculo para integrais de linha

Na Secção 10.11 generalizámos o segundo teorema fundamental do cálculo a integrais de linha. Aqui vamos igualmente generalizar o primeiro teorema fundamental. Lembramos que o primeiro teorema fundamental estabelece que todo o integral indefinido de uma função contínua f tem uma derivada igual a f , isto é, se

$$\varphi(x) = \int_a^x f(t) dt,$$

então nos pontos de continuidade de f temos

$$\varphi'(x) = f(x).$$

Para obtermos a generalização deste teorema aos integrais de linha começamos com um campo vectorial f , contínuo num conjunto conexo aberto S , e integramo-lo ao longo de uma curva C seccionalmente regular entre um ponto fixo a de S a um ponto arbitrário x . Designemos então por φ o campo escalar definido por

$$\varphi(x) = \int_a^x f \cdot d\alpha,$$

onde α é a função que define C . Visto S ser conexo, cada ponto x de S pode ser alcançado por uma tal curva. Para que esta definição de $\varphi(x)$ não seja ambígua, necessitamos saber que o integral depende unicamente de x e não da curva particular considerada para unir a a x . Portanto, é natural exigir que o integral de linha de f seja independente da linha considerada em S . Sob estas considerações, a generalização do primeiro teorema fundamental toma a forma seguinte:

TEOREMA 10.4. PRIMEIRO TEOREMA FUNDAMENTAL PARA INTEGRAIS DE LINHA. *Se f é um campo vectorial contínuo num conjunto conexo S de \mathbb{R}^n , se o integral de linha de f é independente da linha considerada em S e se a é um ponto fixo de S e o campo escalar φ definido em S é dado por*

$$\varphi(x) = \int_a^x f \cdot d\alpha,$$

com α uma qualquer linha seccionalmente regular de S unindo a com x , então o gradiente de φ existe e é igual a f , isto é,

$$\nabla\varphi(x) = f(x) \quad \text{para todo } x \text{ em } S.$$

Demonstração. Provaremos que a derivada parcial $D_k\varphi(x)$ existe e é igual a $f_k(x)$, a componente de ordem k de $f(x)$ para cada $k = 1, 2, \dots, n$ e cada valor de x em S .

Seja $B(x; r)$ uma n -bola com centro em x e raio r , situada em S . Se y é um vector tangente unitário, o ponto $x + hy$ também está em S para todo o real h satisfazendo a $0 < |h| < r$, e podemos formar a razão incremental

$$\frac{\varphi(x + hy) - \varphi(x)}{h}.$$

Devido à propriedade aditiva dos integrais de linha, o numerador deste cociente pode escrever-se

$$\varphi(x + hy) - \varphi(x) = \int_x^{x+hy} f \cdot d\alpha,$$

e a linha unindo x a $x + hy$ pode ser uma qualquer seccionalmente regular contida em S . Em particular, podemos utilizar o segmento de recta definido por

$$\alpha(t) = x + thy, \quad \text{onde } 0 \leq t \leq 1.$$

Porque $\alpha'(t) = h\mathbf{r}$, a razão incremental vem

$$(10.4) \quad \frac{\varphi(\mathbf{x} + h\mathbf{y}) - \varphi(\mathbf{x})}{h} = \int_0^1 f(\mathbf{x} + t\mathbf{y}) \cdot \mathbf{y} \, dt.$$

Tomemos agora $\mathbf{y} = \mathbf{e}_k$, o vector coordenado unitário de ordem k e observemos que o integrando vem $f(\mathbf{x} + t\mathbf{y}) \cdot \mathbf{y} = f_k(\mathbf{x} + t\mathbf{e}_k)$. Fazemos depois a mudança de variável $u = ht$, $du = hdt$ e escrevemos (10.4) na forma

$$(10.5) \quad \frac{\varphi(\mathbf{x} + h\mathbf{e}_k) - \varphi(\mathbf{x})}{h} = \frac{1}{h} \int_0^h f_k(\mathbf{x} + u\mathbf{e}_k) \, du = \frac{g(h) - g(0)}{h},$$

onde g é a função definida no intervalo aberto $(-r, r)$ pela equação

$$g(t) = \int_0^t f_k(\mathbf{x} + u\mathbf{e}_k) \, du.$$

Como cada componente f_k é contínua em S , o primeiro teorema fundamental para integrais definidos diz-nos que $g'(t)$ existe para todo t em $(-r, r)$ e que

$$g'(t) = f_k(\mathbf{x} + t\mathbf{e}_k).$$

Em particular, $g'(0) = f_k(\mathbf{x})$. Portanto, se fizermos $h \rightarrow 0$ em (10.5) encontramos

$$\lim_{h \rightarrow 0} \frac{\varphi(\mathbf{x} + h\mathbf{e}_k) - \varphi(\mathbf{x})}{h} = \lim_{h \rightarrow 0} \frac{g(h) - g(0)}{h} = g'(0) = f_k(\mathbf{x}).$$

Prove isto que a derivada parcial $D_k \varphi(\mathbf{x})$ existe e é igual a $f_k(\mathbf{x})$, como se afirmou.

10.15. Condições necessárias e suficientes para que um campo de vectores seja um gradiente

Os primeiro e segundo teoremas fundamentais para os integrais de linha dizem-nos que uma condição necessária e suficiente para que um campo de vectores contínuos seja um gradiente num conjunto conexo aberto é que o seu integral de linha entre dois pontos quaisquer seja independente da linha considerada. Devemos agora provar que esta condição é equivalente à afirmação de que o integral de linha ao longo de qualquer linha fechada, seccionalmente regular, é zero. Todas estas condições estão reunidas no seguinte teorema.

TEOREMA 10.5. CONDIÇÕES NECESSÁRIAS E SUFICIENTES PARA QUE UM CAMPO VECTORIAL SEJA UM GRADIENTE. Se \mathbf{f} é um campo vectorial contínuo num conjunto conexo aberto S de \mathbb{R}^n , então são equivalentes as três proposições seguintes:

- (a) \mathbf{f} é o gradiente de alguma função potencial em S .
- (b) O integral de linha de \mathbf{f} é independente da linha considerada em S .
- (c) O integral de linha de \mathbf{f} , ao longo de qualquer linha fechada seccionalmente regular contida em S , é zero.

Demonstração. Vamos provar que (b) implica (a), (a) implica (c) e (c) implica (b). A proposição (b) implica (a) devido ao primeiro teorema fundamental. O segundo teorema fundamental mostra que (a) implica (c).

Para completar a demonstração vamos provar que (c) implica (b). Admitamos que é verdadeira a proposição (c) e sejam C_1 e C_2 duas quaisquer curvas seccionalmente regulares em S e com as mesmas extremidades. Seja C_1 o gráfico de uma função α definida num intervalo $[a, b]$, e C_2 o gráfico de β definida em $[c, d]$.

Consideremos uma nova função γ do modo seguinte:

$$\gamma(t) = \begin{cases} \alpha(t) & \text{se } a \leq t \leq b, \\ \beta(b + d - t) & \text{se } b \leq t \leq b + d - c. \end{cases}$$

Então γ define uma curva fechada C tal que

$$\oint_C f \cdot d\gamma = \int_{C_1} f \cdot d\alpha - \int_{C_2} f \cdot d\beta.$$

Visto que $\oint_C f \cdot d\gamma = 0$ devido a (c), temos $\int_{C_1} f \cdot d\alpha = \int_{C_2} f \cdot d\beta$ pelo que o integral de f é independente da linha, o que prova (b). Podemos pois afirmar que (a), (b) e (c) são equivalentes.

Nota: Se $\oint_C f \neq 0$ para uma determinada curva fechada C , então f não é um gradiente. Contudo, ser o integral de linha $\oint_C f$ zero para uma curva fechada particular C ou ainda para infinitamente muitas curvas fechadas, não implica necessariamente que f seja um gradiente. Por exemplo, o leitor pode comprovar facilmente que o integral de linha do campo vectorial $f(x, y) = x\mathbf{i} + xy\mathbf{j}$ é nulo para todo o círculo com centro na origem. Todavia, este campo vectorial não é um gradiente.

10.16. Condições necessárias para que um campo vectorial seja um gradiente

O primeiro teorema fundamental pode ser utilizado para verificar se sim ou não um dado campo vectorial é um gradiente num conjunto conexo aberto S . Se o integral de linha de f é independente da linha considerada em S , definimos muito simplesmente um campo escalar ϕ integrando f desde um ponto fixo até um ponto arbitrário x de S , ao longo uma linha conveniente de S . Calculamos então as derivadas parciais de ϕ e comparamos $D_k\phi$ com f_k , a componente de ordem k de f . Se $D_k\phi(x) = f_k(x)$, para todo x de S e k qualquer, então f é um gradiente em S e ϕ um potencial. Se $D_k\phi(x) \neq f_k(x)$ para algum k e algum x , então f não é um gradiente em S .

O teorema que se segue constitui outro critério para verificar quando um campo vectorial f não é um gradiente. Este critério é particularmente útil na prática devido a não exigir qualquer integração.

TEOREMA 10.6. CONDIÇÕES NECESSÁRIAS PARA QUE UM CAMPO VECTORIAL SEJA UM GRADIENTE. *Seja $f = (f_1, \dots, f_n)$ um campo vectorial continuamente diferenciável*

num conjunto aberto S de \mathbb{R}^n . Se f é um gradiente em S , então as derivadas parciais das componentes de f estão relacionadas pelas fórmulas

$$(10.6) \quad D_i f_j(x) = D_j f_i(x)$$

para $i, j = 1, 2, \dots, n$ e qualquer x de S .

Demonstração. Se f é um gradiente, então $f = \nabla \varphi$ para alguma função potencial φ . Isto significa que

$$f_j = D_j \varphi$$

para cada $j = 1, 2, \dots, n$. Derivando ambos os membros desta igualdade em relação a x_i encontramos

$$D_i f_j = D_i D_j \varphi.$$

Analogamente, temos

$$D_j f_i = D_j D_i \varphi.$$

Mas como as derivadas parciais $D_i f_j$ e $D_j f_i$ são contínuas em S , as duas derivadas parciais mistas $D_i D_j \varphi$ e $D_j D_i \varphi$ são iguais em S e portanto fica demonstrado (10.6).

EXEMPLO 1. Verificar se o campo vectorial

$$f(x, y) = 3x^2 y i + x^3 y j$$

é ou não um gradiente em qualquer subconjunto aberto de \mathbb{R}^2 .

Resolução. Aqui temos

$$f_1(x, y) = 3x^2 y, \quad f_2(x, y) = x^3 y.$$

As derivadas parciais $D_2 f_1$ e $D_1 f_2$ são dadas por

$$D_2 f_1(x, y) = 3x^2, \quad D_1 f_2(x, y) = 3x^2 y.$$

Visto $D_2 f_1(x, y) \neq D_1 f_2(x, y)$ excepto quando $x = 0$ ou $y = 1$, este campo vectorial não é um gradiente em nenhum subconjunto aberto de \mathbb{R}^2 .

O exemplo que se segue mostra que as condições do teorema 10.6 não são sempre suficientes para que um campo vectorial seja um gradiente.

EXEMPLO 2. Seja S o conjunto de todos $(x, y) \neq (0, 0)$ em \mathbb{R}^2 e f um campo vectorial definido em S por

$$f(x, y) = \frac{-y}{x^2 + y^2} \mathbf{i} + \frac{x}{x^2 + y^2} \mathbf{j}.$$

Mostrar que $D_1 f_2 = D_2 f_1$ em todo o conjunto S mas que, apesar disso, f não é um gradiente em S .

Resolução. O leitor pode verificar com facilidade que $D_1 f_2(x, y) = D_2 f_1(x, y)$ para todo (x, y) de S . (Ver Exercício 17, Secção 10.18).

Para provar que f não é um gradiente em S calculamos o integral de linha de f ao longo do círculo unitário definido por

$$\alpha(t) = \cos t \mathbf{i} + \sin t \mathbf{j}, \quad 0 \leq t \leq 2\pi.$$

Obtemos

$$\oint f \cdot d\alpha = \int_0^{2\pi} f[\alpha(t)] \cdot \alpha'(t) dt = \int_0^{2\pi} (\sin^2 t + \cos^2 t) dt = 2\pi.$$

Visto o integral de linha ao longo de um contorno fechado não ser nulo, f não é um gradiente em S . Outras propriedades deste campo vectorial serão estudadas no Exercício 18 da Secção 10.18.

No final do presente capítulo provaremos que as condições necessárias do teorema 10.6 são também suficientes se forem satisfeitas num conjunto *convexo* aberto. (Ver teorema 10.9.)

10.17. Métodos especiais de construção de funções potenciais

O primeiro teorema fundamental para os integrais de linha dá-nos também um método para a construção de funções potenciais. Se f é um gradiente contínuo num conjunto conexo S , o integral de linha de f é independente da linha em S . Deste modo podemos determinar uma função potencial ϕ , integrando f de algum ponto fixo a a um ponto arbitrário x de S , recorrendo a qualquer linha seccionalmente regular situada em S . O campo escalar assim obtido depende da escolha do ponto inicial a . Se partirmos doutro ponto inicial, por exemplo b , obtemos uma nova função potencial ψ . Mas, devido à propriedade aditiva dos integrais de linha, ϕ e ψ podem diferir unicamente por uma constante, sendo esta constante o valor do integral de f de a a b .

Os exemplos que se seguem ilustram o uso de diferentes escolhas das linhas de integração.

EXEMPLO 1. Construção de um potencial num rectângulo aberto. Se f é um gradiente contínuo num rectângulo aberto de \mathbb{R}^n , pode construir-se um potencial ϕ integrando de um ponto fixo até um ponto arbitrário ao longo de linhas formadas por segmentos paralelos aos eixos coordenados. Um exemplo bidimensional está representado na figura 10.5. Podemos integrar em primeiro lugar de (a, b) até (x, b) ao longo do segmento paralelo a OX , e em seguida de (x, b) até (x, y) ao longo de um segmento paralelo a OY . Ao longo do segmento paralelo a OX a representação paramétrica

$$\alpha(t) = ti + bj, \quad a \leq t \leq x,$$

e ao longo do segmento paralelo a OY recorreremos à representação

$$\alpha(t) = xi + tj, \quad b \leq t \leq y.$$

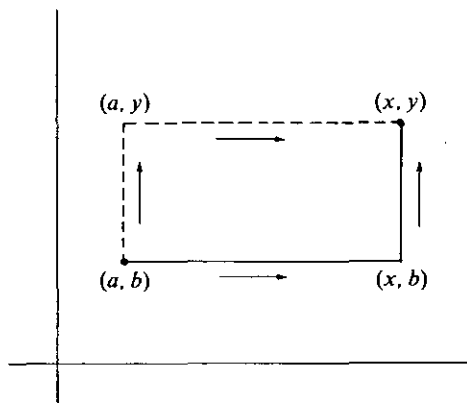


FIG. 10.5. Duas linhas poligonais de (a, b) a (x, y) .

Se $f(x, y) = f_1(x, y)i + f_2(x, y)j$, a fórmula resultante para um potencial $\phi(x, y)$ é

$$(10.7) \quad \phi(x, y) = \int_a^x f_1(t, b) dt + \int_b^y f_2(x, t) dt.$$

Podêríamos também integrar em primeiro lugar de (a, b) a (a, y) ao longo do segmento paralelo a OY e depois de (a, y) até (x, y) ao longo de um segmento paralelo a OX representado a tracejado na figura 10.5. Isto dá-nos uma outra fórmula para $\phi(x, y)$.

$$(10.8) \quad \phi(x, y) = \int_b^y f_2(a, t) dt + \int_a^x f_1(t, y) dt.$$

Ambas as fórmulas (10.7) e (10.8) conduzem ao mesmo valor para $\phi(x, y)$ devido ao integral de linha de um gradiente ser independente da linha.

EXEMPLO 2. *Construção de uma função potencial utilizando integrais indefinidos.* O uso de integrais indefinidos simplificam-nos, muitas vezes, o cálculo de funções potenciais. Por exemplo, suponhamos que um campo vectorial tridimensional $f = (f_1, f_2,$

f_3 é o gradiente de uma função potencial φ num conjunto aberto S de \mathbb{R}^3 . Então tem-se

$$\frac{\partial \varphi}{\partial x} = f_1, \quad \frac{\partial \varphi}{\partial y} = f_2, \quad \frac{\partial \varphi}{\partial z} = f_3$$

em todo o conjunto S . Recorrendo aos integrais indefinidos e integrando a primeira destas equações relativamente a x (mantendo y e z constantes) encontramos

$$\varphi(x, y, z) = \int f_1(x, y, z) dx + A(y, z),$$

onde $A(y, z)$ é uma “constante de integração” a determinar. Analogamente, se integramos a equação $\partial \varphi / \partial y = f_2$ relativamente a y e $\partial \varphi / \partial z = f_3$ relativamente a z obtemos mais duas relações

$$\varphi(x, y, z) = \int f_2(x, y, z) dy + B(x, z)$$

e

$$\varphi(x, y, z) = \int f_3(x, y, z) dz + C(x, y),$$

onde $B(x, z)$ e $C(x, y)$ são funções a determinar. Determinar φ significa determinar três funções $A(y, z)$, $B(x, z)$ e $C(x, y)$ tais que as três expressões de $\varphi(x, y, z)$ tenham segundos membros equivalentes. Em muitos casos isso pode conseguir-se por análise directa, como se verifica no exemplo que a seguir se apresenta.

EXEMPLO 3. Determinar uma função potencial φ para o campo vectorial definido em \mathbb{R}^3 pela equação

$$f(x, y, z) = (2xyz + z^2 - 2y^2 + 1)\mathbf{i} + (x^2z - 4xy)\mathbf{j} + (x^2y + 2xz - 2)\mathbf{k}.$$

Resolução. Sem saber *a priori* se f admite ou não uma função potencial φ , vamos tentar determinar um potencial como foi referido no Exemplo 2, admitindo que tal potencial φ existe.

Integrando a componente f_1 relativamente a x encontramos

$$\varphi(x, y, z) = \int (2xyz + z^2 - 2y^2 + 1) dx + A(y, z) = x^2yz + xz^2 - 2xy^2 + x + A(y, z).$$

Integrando f_2 em relação a y , e depois f_3 em relação a z , encontramos

$$\varphi(x, y, z) = \int (x^2z - 4xy) dy + B(x, z) = x^2yz - 2xy^2 + B(x, z),$$

$$\varphi(x, y, z) = \int (x^2y + 2xz - 2) dz + C(x, y) = x^2yz + xz^2 - 2z + C(x, y).$$

Por simples análise das expressões obtidas, vemos que a escolha de $A(y, z) = -2z$, $B(x, z) = xz^2 + x - 2z$, e $C(x, y) = x - 2xy^2$ fará com que as três expressões coincidam; por conseguinte φ , dada por

$$\varphi(x, y, z) = x^2yz + xz^2 - 2xy^2 + x - 2z$$

é um potencial correspondente a f em \mathbb{R}^3 .

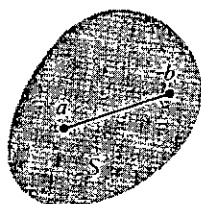
EXEMPLO 4. Construção de um potencial num conjunto convexo. Um conjunto S de \mathbb{R}^n diz-se *convexo* se cada par de pontos de S pode unir-se por um segmento de recta, cujos pontos pertencem todos a S . Na figura 10.6 apresenta-se um exemplo. Todo o conjunto convexo é conexo.

Se f é um gradiente contínuo num conjunto aberto convexo, então pode construir-se uma função potencial φ por integração de f de um ponto fixo a de S até um ponto arbitrário x , ao longo do segmento de recta unindo a a x . Tal segmento pode definir-se parametricamente pela função

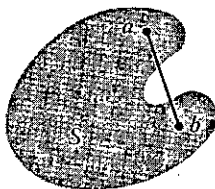
$$\alpha(t) = a + t(x - a), \quad \text{com} \quad 0 \leq t \leq 1.$$

Daqui resulta $\alpha'(t) = x - a$, pelo que o potencial correspondente é dado pelo integral

$$(10.9) \quad \varphi(x) = \int_0^1 f(a + t(x - a)) \cdot (x - a) dt.$$



Convexo



Não convexo

FIG. 10.6. Num conjunto convexo S , o segmento unindo dois pontos quaisquer a e b de S está em S .

Se S contém a origem podemos tomar $a = O$ escrever (10.9) na forma mais simples

$$(10.10) \quad \varphi(x) = \int_0^1 f(tx) \cdot x dt.$$

10.18. Exercícios

Em cada um dos Exercícios 1 a 12, define-se um campo vectorial f , pelas fórmulas dadas. Para cada um determinar se f é ou não o gradiente de um campo escalar. Quando f for um gradiente, determinar a correspondente função potencial φ .

1. $f(x, y) = xi + yj$.
 2. $f(x, y) = 3x^2yi + x^3j$.
 3. $f(x, y) = (2xe^y + y)i + (x^2e^y + x - 2y)j$.
 4. $f(x, y) = (\sin y - y \sin x + x)i + (\cos x + x \cos y + y)j$.
 5. $f(x, y) = [\sin(xy) + xy \cos(xy)]i + x^2 \cos(xy)j$.
 6. $f(x, y, z) = xi + yj + zk$.
 7. $f(x, y, z) = (x + z)i - (y + z)j + (x - y)k$.
 8. $f(x, y, z) = 2xy^3i + x^2z^3j + 3x^2yz^2k$.
 9. $f(x, y, z) = 3y^4z^2i + 4x^3z^2j - 3x^2y^2k$.
 10. $f(x, y, z) = (2x^2 + 8xy^2)i + (3x^3y - 3xy)j - (4y^2z^2 + 2x^3z)k$.
 11. $f(x, y, z) = (y^2 \cos x + z^3)i - (4 - 2y \sin x)j + (3xz^2 + 2)k$.
 12. $f(x, y, z) = (4xy - 3x^2z^2 + 1)i + 2(x^2 + 1)j - (2xz^2 + 3z^2)k$.
13. Um fluido desloca-se no plano XOY , movendo-se cada partícula sobre uma recta que passa pela origem. Se uma partícula estiver a uma distância r da origem a sua velocidade é ar^n , com a e n constantes.
- (a) Determinar os valores de a e n para os quais o campo vectorial das velocidades é o gradiente de um certo campo escalar.
 - (b) Determinar uma função potencial da velocidade sempre que esta seja um gradiente. O caso em que $n = -1$ deve ser considerado separadamente.
14. Se ambos φ e ψ são funções potencial para um campo vectorial contínuo f num conjunto conexo aberto S de \mathbb{R}^n , provar que $\varphi - \psi$ é constante em S .
15. Seja S o conjunto de todos os pontos $x \neq O$ em \mathbb{R}^n . Sejam $r = \|x\|$ e f o campo vectorial definido em S por

$$f(x) = r^p x,$$

com p uma constante real. Determinar uma função potencial para f em S . O caso em que $p = -2$ deve ser tratado separadamente.

16. Resolver o Exercício 15 para o campo vectorial definido pela equação

$$f(x) = \frac{g'(r)}{r} x,$$

com g uma função real admitindo derivada contínua em todo \mathbb{R}^1 .

Os exercícios que se seguem referem-se ao campo vectorial f definido no conjunto S de todos os pontos $(x, y) \neq (0, 0)$ em \mathbb{R}^2 pela equação

$$f(x, y) = -\frac{y}{x^2 + y^2}i + \frac{x}{x^2 + y^2}j.$$

No Exemplo 2 da Secção 10.16 mostrou-se que f não é um gradiente em S , muito embora $D_1 f_2 = D_2 f_1$ por todo S .

17. Verificar que para todo o ponto (x, y) de S se tem

$$D_1 f_2(x, y) = D_2 f_1(x, y) = \frac{y^2 - x^2}{(x^2 + y^2)^2}.$$

18. Este exercício mostra que f é um gradiente no conjunto

$$T = \mathbb{R}^2 - \{(x, y) \mid y = 0, x \leq 0\},$$

formado por todos os pontos de XOY excepto os do semi-eixo negativo OX .

(a) Se $(x, y) \in T$, exprimir x e y em coordenadas polares,

$$x = r \cos \theta, \quad y = r \sin \theta,$$

onde $r > 0$ e $-\pi < \theta < \pi$. Provar que θ é dado pelas fórmulas

$$\theta = \begin{cases} \arctan \frac{y}{x} & \text{se } x > 0, \\ \frac{\pi}{2} & \text{se } x = 0, \\ \arctan \frac{y}{x} + \pi & \text{se } x < 0. \end{cases}$$

[Deve ter-se presente a definição da função arco tangente: Para qualquer real t , $\arctan t$ é o número real único φ que verifica as duas condições $\tan \varphi = t$ e $-\pi/2 < \varphi < \pi/2$.]

(b) Provar que

$$\frac{\partial \theta}{\partial x} = -\frac{y}{x^2 + y^2}, \quad \frac{\partial \theta}{\partial y} = \frac{x}{x^2 + y^2}$$

para todo (x, y) em T . Isto prova que θ é uma função potencial para f no conjunto T .

Este exercício ilustra o facto de que a existência de uma função potencial depende não somente do campo vectorial f , mas também do conjunto em que o campo vectorial se define.

10.19. Aplicações às equações diferenciais exactas de primeira ordem

Algumas equações diferenciais de primeira ordem podem resolver-se com recurso às funções potenciais. Suponhamos uma equação diferencial de primeira ordem da forma

$$y' = f(x, y).$$

Se multiplicarmos ambos os membros por um factor não nulo, $Q(x, y)$, transformamos esta equação em outra da forma $Q(x, y)y' - f(x, y)Q(x, y) = 0$. Escrevendo $P(x, y)$ em vez de $-f(x, y)Q(x, y)$ e usando a notação de Leibniz para as derivadas, escrevendo dy/dx em vez de y' , a equação diferencial toma a forma

$$(10.11) \quad P(x, y) dx + Q(x, y) dy = 0.$$

Sapomos que P e Q são contínuas em algum conjunto aberto conexo S do plano. A cada tal equação diferencial podemos associar um campo vectorial V , da forma

$$V(x, y) = P(x, y)i + Q(x, y)j.$$

As componentes P e Q são os coeficientes de dx e dy na equação (10.11). A equação diferencial (10.11) diz-se *exacta* em S se o campo vectorial V for o gradiente de um potencial, isto é, se $V(x, y) = \nabla \phi(x, y)$ para todo ponto (x, y) de S , onde ϕ é um certo campo escalar. Quando uma tal função ϕ existe, tem-se $\partial \phi / \partial x = P$ e $\partial \phi / \partial y = Q$, e a equação diferencial (10.11) escreve-se

$$\frac{\partial \phi}{\partial x} dx + \frac{\partial \phi}{\partial y} dy = 0.$$

Vamos demonstrar agora que toda a solução desta equação diferencial verifica a relação $\phi(x, y) = C$, sendo C constante. Mais precisamente, supomos que existe uma solução Y da equação diferencial (10.11) definida num intervalo aberto (a, b) , tal que o ponto $(x, Y(x))$ esteja em S para todo x de (a, b) . Vamos provar que

$$\phi[x, Y(x)] = C$$

para uma certa constante C . Com tal finalidade introduzimos a função composta g definida em (a, b) pela equação

$$g(x) = \phi[x, Y(x)].$$

A derivada de g vem então

$$(10.12) \quad g'(x) = D_1 \phi[x, Y(x)] + D_2 \phi[x, Y(x)] Y'(x) = P(x, y) + Q(x, y) y',$$

onde $y = Y(x)$ e $y' = Y'(x)$. Se y satisfaz a (10.11), então $P(x, y) + Q(x, y) y' = 0$, pelo que $g'(x) = 0$ para todo x em (a, b) e, consequentemente, g é constante em (a, b) . Isto prova que toda a solução y satisfaz à equação $\phi(x, y) = C$.

Podemos agora conduzir a raciocínio em sentido inverso para obtermos uma solução da equação diferencial. Suponhamos que a equação

$$(10.13) \quad \phi(x, y) = C$$

define y como uma função derivável de x , por exemplo $y = Y(x)$ para x um intervalo (a, b) , e seja $g(x) = \phi[x, Y(x)]$. A equação (10.13) implica que g é constante em (a, b) . Logo, por (10.12), $P(x, y) + Q(x, y) y' = 0$, de maneira que y é uma solução. Provamos assim o seguinte teorema:

TEOREMA 10.7. *Admita-se que a equação diferencial*

$$(10.14) \quad P(x, y) dx + Q(x, y) dy = 0$$

é exacta num conjunto aberto conexo S , e seja φ um campo escalar satisfazendo a

$$\frac{\partial \varphi}{\partial x} = P \quad e \quad \frac{\partial \varphi}{\partial y} = Q$$

em todo S . Então toda a solução $y = Y(x)$ de (10.14), cujo gráfico esteja em S , satisfaz à equação $\varphi[x, Y(x)] = C$ para certa constante C . Inversamente, se a equação

$$\varphi(x, y) = C$$

define y implicitamente como uma função derivável de x , então esta função é uma solução da equação diferencial (10.14).

O teorema precedente proporciona-nos um método directo de resolução das equações diferenciais exactas de primeira ordem. Construímos uma função potencial φ e escrevemos $\varphi(x, y) = C$, com C constante. Sempre que esta equação potencial φ é tamente como uma função de x , a correspondente y satisfaz a (10.14). Portanto podemos usar a equação (10.13) como uma representação de uma família a um parâmetro de curvas integrais. Evidentemente que os únicos valores admissíveis de C são aqueles para os quais $\varphi(x_0, y_0) = C$ para algum (x_0, y_0) de S .

EXEMPLO 1. Seja dada a equação diferencial

$$\frac{dy}{dx} = -\frac{3x^2 + 6xy^2}{6x^2y + 4y^3}.$$

Desembaraçando de denominadores, podemos escrever a equação

$$(3x^2 + 6xy^2) dx + (6x^2y + 4y^3) dy = 0.$$

Este é agora um caso particular de (10.14), com $P(x, y) = 3x^2 + 6xy^2$ e $Q(x, y) = 6x^2y + 4y^3$. Integrando P relativamente a x e Q relativamente a y e comparando os resultados, verificamos que a função potencial vem dada por

$$\varphi(x, y) = x^3 + 3x^2y^2 + y^4.$$

Pelo teorema 10.7, cada solução da equação diferencial satisfaz a

$$x^3 + 3x^2y^2 + y^4 = C$$

para algum C . Esta representa, na forma implícita, uma família de curvas integrais. Neste caso particular a equação é do segundo grau em y^2 e pode resolver-se de modo a obter-se uma fórmula dando y em função de x e C .

EXEMPLO 2. Considere-se a equação diferencial de primeira ordem

$$(10.15) \quad y \, dx + 2x \, dy = 0.$$

Neste caso $P(x, y) = y$ e $Q(x, y) = 2x$. Visto que $\partial P / \partial y = 1$ e $\partial Q / \partial x = 2$, esta equação diferencial não é exacta. Contudo se multiplicamos ambos os membros por y , obtemos uma equação diferencial que já é exacta, a saber

$$(10.16) \quad y^2 \, dx + 2xy \, dy = 0.$$

Um potencial para o campo vectorial $y^2 \mathbf{i} + 2xy \mathbf{j}$ é $\phi(xy) = xy^2$, e toda a solução de (10.16) verifica a relação $xy^2 = C$. Esta relação representa também uma família de curvas integrais para a Equação (10.15).

O factor y , que transformou (10.15) numa equação diferencial exacta, chama-se um *factor integrante*. Em geral, se a multiplicação de uma equação diferencial linear de primeira ordem por um factor não nulo $\mu(x, y)$ a transforme numa equação diferencial exacta, o factor $\mu(x, y)$ diz-se um factor integrante da equação original. Uma equação diferencial pode admitir mais do que um factor integrante. Por exemplo, $\mu(x, y) = 2xy^3$ é outro factor integrante de (10.15). Nos exercícios que se seguem analisam-se alguns casos particulares de equações diferenciais para as quais se podem encontrar facilmente factores integrantes.

10.20. Exercícios

Provar que as equações dos Exercícios 1 a 5 são exactas, e para cada uma determinar uma família, a um parâmetro, de curvas integrais.

1. $(x + 2y) \, dx + (2x + y) \, dy = 0$.
2. $2xy \, dx + x^2 \, dy = 0$.
3. $(x^2 - y) \, dx - (x + \sinh^2 y) \, dy = 0$.
4. $4 \sin x \sin 3y \cos x \, dx - 3 \cos 3y \cos 2x \, dy = 0$.
5. $(3x^2y + 8xy^2) \, dx + (x^3 + 8x^2y + 12ye^y) \, dy = 0$.
6. Provar que a equação diferencial linear de primeira ordem, $y' + P(x)y = Q(x)$, admite o factor integrante $\mu(x) = e^{\int P(x) \, dx}$. Resolver a equação neste caso.
7. Seja $\mu(x, y)$ um factor integrante da equação diferencial $P(x, y) \, dx + Q(x, y) \, dy = 0$. Provar que

$$\frac{\partial P}{\partial y} - \frac{\partial Q}{\partial x} = Q \frac{\partial}{\partial x} \log |\mu| - P \frac{\partial}{\partial y} \log |\mu|.$$

Usar esta equação para deduzir as seguintes regras de determinação de factores integrantes:

(a) Se $(\partial P / \partial y - \partial Q / \partial x) / Q$ é unicamente uma função de x , por exemplo $f(x)$, então $e^{\int f(x) \, dx}$ é um factor integrante.

(b) Se $(\partial Q / \partial x - \partial P / \partial y) / P$ é unicamente função de y , por exemplo $g(y)$, então $e^{\int g(y) \, dy}$ é um factor integrante.

8. Aplicar o Exercício 7 na determinação de factores integrantes para as seguintes equações, e determinar as respectivas famílias de curvas integrais a um parâmetro.

$$(a) y dx - (2x + y) dy = 0.$$

$$(b) (x^3 + y^3) dx - xy^2 dy = 0.$$

9. Se $\partial P/\partial y - \partial Q/\partial x = f(x)Q(x, y) - g(y)P(x, y)$, provar que $\exp\{\int f(x)dx + \int g(y)dy\}$ é um factor integrante da equação diferencial $P(x, y)dx + Q(x, y)dy = 0$. Determinar o correspondente factor integrante de cada uma das equações seguintes e obter uma família a um parâmetro de curvas integrais.

$$(a) (2x^2y + y^2) dx + (2x^3 - xy) dy = 0.$$

$$(b) (e^x \sec y - \tan y) dx + dy = 0.$$

10. As equações diferenciais seguintes têm um factor integrante comum. Determinar tal factor integrante e obter uma família a um parâmetro, de curvas integrais para cada equação

$$(3y + 4xy^2) dx + (4x + 5x^2y) dy = 0,$$

$$(6y + x^2y^2) dx + (8x + x^3y) dy = 0.$$

10.21. Funções potenciais em conjuntos convexos

No teorema 10.6 provou-se que as condições

$$D_i f_j(x) = D_j f_i(x)$$

são *necessárias* para que um campo vectorial $f = (f_1, f_2, \dots, f_n)$, continuamente diferenciável, seja um gradiente num conjunto aberto S de \mathbf{R}^n . Demonstrou-se aí, por meio de um exemplo, que estas condições nem sempre são suficientes. Nesta secção vamos provar que as condições são suficientes sempre que o conjunto S seja convexo. A demonstração fará uso do seguinte teorema relativo à derivação sob o sinal de integral.

TEOREMA 10.8. *Seja S um intervalo fechado em \mathbf{R}^n com interior não vazio e $J = [a, b]$ um intervalo fechado em \mathbf{R}^1 . Seja J_{n+1} o intervalo fechado $S \times J$ em \mathbf{R}^{n+1} . Represente-se cada ponto de J_{n+1} por (x, t) , em que $x \in S$ e $t \in J$.*

$$(x, t) = (x_1, \dots, x_n, t).$$

Admita-se que ϕ é um campo escalar definido em J_{n+1} , tal que a derivada parcial $D_k \phi$ seja contínua em J_{n+1} , com k um dos valores $1, 2, \dots, n$. Defina-se um campo escalar φ em S pela equação

$$\varphi(x) = \int_a^b \psi(x, t) dt.$$

A derivada parcial $D_k \phi$ existe em cada ponto interior de S e é dada pela fórmula

$$D_k \varphi(x) = \int_a^b D_k \psi(x, t) dt.$$

Quer dizer, tem-se

$$D_k \int_a^b \psi(x, t) dt = \int_a^b D_k \psi(x, t) dt.$$

Nota: Este teorema indica que se pode derivar sob o sinal de integral.

Demonstração. Escolhamos qualquer x do interior de S . Visto que $\text{int } S$ é aberto, existe um $r > 0$ tal que $x + he_k \in \text{int } S$ para todo h satisfazendo a $0 < h < r$. Aqui e_k é o vector coordenado unitário de ordem k em \mathbb{R}^n . Para tal h temos

$$(10.17) \quad \varphi(x + he_k) - \varphi(x) = \int_a^b \{\psi(x + he_k, t) - \psi(x, t)\} dt.$$

Aplicando o teorema da média ao integrando temos

$$\psi(x + he_k, t) - \psi(x, t) = h D_k \psi(z, t),$$

com z pertencente ao segmento de recta que une x com $x + he_k$. Deste modo (10.17) vem

$$\frac{\varphi(x + he_k) - \varphi(x)}{h} = \int_a^b D_k \psi(z, t) dt.$$

Portanto

$$\frac{\varphi(x + he_k) - \varphi(x)}{h} - \int_a^b D_k \psi(x, t) dt = \int_a^b \{D_k \psi(z, t) - D_k \psi(x, t)\} dt.$$

O último integral tem um modo absoluto que não excede

$$\int_a^b |D_k \psi(z, t) - D_k \psi(x, t)| dt \leq (b - a) \max |D_k \psi(z, t) - D_k \psi(x, t)|$$

onde o máximo é alcançado para todo z do segmento que une x com $x + he_k$, e todo t de $[a, b]$. Tendo em conta a continuidade uniforme de D_k em $S \times J$ (teorema 9.10) chegamos à conclusão de que para todo $\epsilon > 0$ existe um $\delta > 0$ tal que esse máximo seja $< \epsilon/(b - a)$, sempre que $0 < |h| < \delta$. Portanto

$$\left| \frac{\varphi(x + he_k) - \varphi(x)}{h} - \int_a^b D_k \psi(x, t) dt \right| < \epsilon \quad \text{sempre que } 0 < |h| < \delta.$$

Isto prova que $D_k \varphi(x)$ existe e é igual a $\int_a^b D_k \psi(x, t) dt$, como se pretendia demonstrar.

Vamos em seguida usar este teorema para dar a seguinte condição necessária e suficiente para que um campo vectorial seja um gradiente num conjunto convexo.

TEOREMA 10.9. *Seja $f = (f_1, \dots, f_n)$ um campo vectorial continuamente diferenciável num conjunto convexo aberto S de \mathbb{R}^n . Então f é um gradiente em S se e só se*

$$(10.18) \quad D_k f_j(x) = D_j f_k(x)$$

para todo x de S e todos os valores de $k, j = 1, 2, \dots, n$.

Demonstração. Sabemos, do teorema 10.6, que a condição é necessária. Para provar que é suficiente, admitimos (10.18) e construímos um potencial φ em S .

Por comodidade, admitimos que S contém a origem. Seja $\varphi(x)$ o integral de f ao longo do segmento de recta de O até um ponto arbitrário x de S . Como se provou atrás, equação (10.10), tem-se

$$\varphi(x) = \int_0^1 f(tx) \cdot x \, dt = \int_0^1 \psi(x, t) \, dt,$$

onde $\psi(x, t) = f(tx) \cdot x$. Existe um subintervalo fechado n -dimensional T de S com interior não vazio tal que ψ satisfaz à hipótese do teorema 10.8 em $T \times J$, com $J = [0, 1]$. Por conseguinte a derivada parcial $D_k \varphi(x)$ existe para cada $k = 1, 2, \dots, n$ e pode ser calculada derivando sob o sinal de integral,

$$D_k \varphi(x) = \int_0^1 D_k \psi(x, t) \, dt.$$

Para calcular $D_k \psi(x, t)$, derivamos o produto escalar $f(tx) \cdot x$ e obtemos

$$\begin{aligned} D_k \psi(x, t) &= f(tx) \cdot D_k x + D_k \{f(tx)\} \cdot x \\ &= f(tx) \cdot e_k + t(D_k f_1(tx), \dots, D_k f_n(tx)) \cdot x \\ &= f_k(tx) + t(D_1 f_k(tx), \dots, D_n f_k(tx)) \cdot x, \end{aligned}$$

sendo-se utilizado, no último passo da demonstração, a relação (10.18). Temos portanto

$$D_k \psi(x, t) = f_k(tx) + t \nabla f_k(tx) \cdot x.$$

Façamos agora $g(t) = f_k(tx)$. Então

$$g'(t) = \nabla f_k(tx) \cdot x$$

de modo que a última fórmula para $D_k \psi(x, t)$ se escreve

$$D_k \psi(x, t) = g(t) + t g'(t).$$

Integrando-a de 0 a 1 obtemos

$$(10.19) \quad D_k \varphi(x) = \int_0^1 D_k \psi(x, t) dt = \int_0^1 g(t) dt + \int_0^1 t g'(t) dt.$$

Calculando o último integral mediante uma integração por partes vem

$$\int_0^1 t g'(t) dt = t g(t) \Big|_0^1 - \int_0^1 g(t) dt = g(1) - \int_0^1 g(t) dt.$$

Consequentemente (10.19) toma a forma

$$D_k \varphi(x) = g(1) = f_k(x).$$

Isto mostra que $\nabla \varphi = f$ em S , o que completa a demonstração.

INTEGRAIS MÚLTIPLOS

11.1. Introdução

No Volume I estudaram-se os integrais de forma $\int_a^b f(x)dx$, primeiramente para funções definidas e limitadas em intervalos finitos, e mais tarde para funções não limitadas e intervalos infinitos. No Capítulo 10 do Volume II generalizou-se o conceito de integral introduzindo-se os integrais de linha. Neste Capítulo generaliza-se o conceito ainda noutra direcção. O intervalo unidimensional $[a, b]$ é substituído por um conjunto bidimensional Q , chamado a *região de integração*. Começamos por considerar regiões rectangulares; mais adiante consideraremos regiões mais gerais com contornos curvilíneos. O integrando é um campo escalar f definido e limitado em Q . O integral resultante diz-se um *integral duplo* e representa-se pelo símbolo

$$\iint_Q f, \quad \text{ou por} \quad \iint_Q f(x, y) dx dy.$$

Tal como no caso unidimensional, os símbolos dx e dy não desempenham qualquer papel na definição do integral duplo; contudo, são úteis no cálculo e transformações de integrais.

O programa a desenvolver neste capítulo consiste de várias etapas. Em primeiro lugar analisaremos a definição de integral duplo. O método é análogo ao caso unidimensional tratado no Volume I. O integral define-se em primeiro lugar para funções em escada e só depois para funções mais gerais. Como no caso unidimensional, a definição não proporciona um método útil para o cálculo de integrais. Veremos que grande número de integrais duplos que se apresentam na prática podem calcular-se por integração unidimensional reiterada. Encontraremos também uma conexão entre integrais duplos e integrais de linha. Apresentam-se depois aplicações dos integrais duplos a problemas de cálculo de áreas, volumes, massas, centro de massa e outros conceitos

Como no caso unidimensional, o valor do integral não se altera se a partição P for substituída por outra partição mais fina P' . Assim, o valor do integral é independente da escolha de P , desde que f seja constante nos subrectângulos abertos de Q .

Por uma questão de brevidade, escrevemos algumas vezes Δx_i em vez de $(x_i - x_{i-1})$ e Δy_j no lugar de $(y_j - y_{j-1})$, vindo para a soma (11.1)

$$\sum_{i=1}^n \sum_{j=1}^m c_{ij} \Delta x_i \Delta y_j.$$

Para termos presente como esta soma é construída, escrevemos para o integral o símbolo

$$\iint_Q f(x, y) dx dy.$$

o qual não é mais do que uma alternativa para a notação $\iint_Q f$.

Note-se que se f é constante no interior de Q , seja $f(x, y) = k$ quando $a < x < b$ e $c < y < d$, tem-se

$$(11.2) \quad \iint_Q f = k(b-a)(d-c),$$

independentemente dos valores de f sobre os lados de Q . Porque se tem

$$b-a = \int_a^b dx \quad \text{e} \quad d-c = \int_c^d dy,$$

a fórmula (11.2) pode de novo escrever-se

$$(11.3) \quad \iint_Q f = \int_c^d \left[\int_a^b f(x, y) dx \right] dy = \int_a^b \left[\int_c^d f(x, y) dy \right] dx.$$

Os integrais figurando no segundo e terceiro membros da igualdade são integrais unidimensionais, e a fórmula diz-se proporcionar um cálculo do integral duplo por integração *repetida* ou *iterada*. Em particular, quando f é uma função em escada do tipo acabado de descrever, podemos escrever

$$\iint_{Q_{ij}} f = \int_{y_{j-1}}^{y_j} \left[\int_{x_{i-1}}^{x_i} f(x, y) dx \right] dy = \int_{x_{i-1}}^{x_i} \left[\int_{y_{j-1}}^{y_j} f(x, y) dy \right] dx.$$

Somando para i e j e considerando (11.1), encontramos que (11.3) é válida para funções em escada.

As propriedades que se apresentam a seguir para os integrais duplos de uma função em escada são generalizações dos correspondentes teoremas para o caso unidimensional. Podem demonstrar-se como consequências directas da definição (11.1), ou recorrendo ao uso da fórmula (11.3) e aos teoremas correspondentes para o caso de uma dimensão. Nos teoremas que se seguem, s e t apresentam funções em escada

definidas num rectângulo Q . Para evitar cair-se em casos particulares triviais supomos que Q é um rectângulo não degenerado, isto é, Q não é nem um ponto nem um segmento de recta.

TEOREMA 11.1. PROPRIEDADE DE LINEARIDADE. *Quaisquer que sejam os números reais c_1 e c_2 tem-se*

$$\iint_Q [c_1 s(x, y) + c_2 t(x, y)] dx dy = c_1 \iint_Q s(x, y) dx dy + c_2 \iint_Q t(x, y) dx dy.$$

TEOREMA 11.2. PROPRIEDADE DA ADITIVIDADE. *Se Q se divide em dois rectângulos Q_1 e Q_2 então*

$$\iint_Q s(x, y) dx dy = \iint_{Q_1} s(x, y) dx dy + \iint_{Q_2} s(x, y) dx dy.$$

TEOREMA 11.3. TEOREMA DE COMPARAÇÃO. *Se $s(x, y) \leq t(x, y)$ para todo (x, y) em Q , tem-se*

$$\iint_Q s(x, y) dx dy \leq \iint_Q t(x, y) dx dy.$$

Em particular, se $t(x, y) \geq 0$ para cada (x, y) em Q , então

$$\iint_Q t(x, y) dx dy \geq 0.$$

As demonstrações destes teoremas são deixadas ao leitor como exercícios.

11.4. A definição de integral duplo de uma função definida e limitada num rectângulo

Seja f uma função que é definida e limitada num rectângulo Q ; concretamente, suponhamos que

$$|f(x, y)| \leq M \quad \text{se } (x, y) \in Q.$$

Então f pode permanecer limitada por cima e por baixo por intermédio de duas funções em escada constantes s e t , onde $s(x, y) = -M$ e $t(x, y) = M$ para todo (x, y) em Q . Consideremos agora duas *quaisquer* funções em escada s e t , definidas em Q , tais que

$$(11.4) \quad s(x, y) \leq f(x, y) \leq t(x, y) \quad \text{para cada ponto } (x, y) \text{ em } Q.$$

DEFINIÇÃO DO INTEGRAL DE UMA FUNÇÃO LIMITADA ESTENDIDA A UM RECTÂNGULO. *Se existir um e um só número I tal que*

$$(11.5) \quad \iint_Q s \leq I \leq \iint_Q t$$

para cada par de funções em escada satisfazendo às desigualdades (11.4), este número I chama-se o integral duplo de f estendido a Q e representa-se pelo símbolo

$$\iint_Q f \quad \text{ou} \quad \iint_Q f(x, y) \, dx \, dy.$$

Quando um tal I existe, a função f diz-se integrável em Q .

11.5. Integrais duplos superior e inferior

A definição do integral duplo é inteiramente análoga ao caso unidimensional. Integrais duplos superiores e inferiores podem também ser definidos em analogia com o caso unidimensional.

Admitamos que f é limitado num rectângulo Q e sejam s e t duas funções em escada satisfazendo a (11.4). Dizemos que s está *abaixo* e t *acima* de f , e escrevemos $s \leq f \leq t$. Seja S o conjunto de todos os números $\iint_Q s$ obtidos quando se toma como s todas as funções em escada abaixo de f , e seja T o conjunto de todos os números $\iint_Q t$ obtido quando t representa cada uma das funções em escada acima de f . Ambos os conjuntos S e T são não vazios, uma vez que f é limitada. Também, $\iint_Q s \leq \iint_Q t$ se $s \leq f \leq t$, pelo que cada elemento de S é menor do que qualquer elemento de T . Portanto S tem um supremo e T um ínfimo e satisfazem às desigualdades

$$\iint_Q s \leq \sup S \leq \inf T \leq \iint_Q t$$

para todo s e t satisfazendo a $s \leq f \leq t$. Isto mostra que ambos os valores $\sup S$ e $\inf T$ satisfazem a (11.5). Consequentemente, f é integrável em Q se e só se $\sup S = \inf T$, caso em que se terá

$$\iint_Q f = \sup S = \inf T.$$

O número $\sup S$ diz-se o *integral inferior* de f e representa-se por $I(f)$. O número $\inf T$ chama-se o *integral superior* de f e representa-se por $\bar{I}(f)$. Então, tem-se

$$I(f) = \sup \left\{ \iint_Q s \mid s \leq f \right\}, \quad \bar{I}(f) = \inf \left\{ \iint_Q t \mid f \leq t \right\}.$$

O precedente argumento prova o seguinte teorema.

TEOREMA 11.4. *Toda a função f limitada num rectângulo Q tem um integral inferior $I(f)$ e um integral superior $\bar{I}(f)$ satisfazendo às desigualdades*

$$\iint_Q s \leq I(f) \leq I(f) \leq \iint_Q t$$

para todas as funções em escada s e t com $s \leq f \leq t$. A função f é integrável em Q se e só se os seus integrais superior e inferior são iguais, caso em que se terá

$$\iint_Q f = I(f) = I(f).$$

Uma vez que as precedentes definições são inteiramente análogas ao caso unidimensional, não deve surpreender que as propriedades de linearidade e aditividade e o teorema de comparação, estabelecidos para funções em escada na Secção 11.3, sejam igualmente válidos para os integrais duplos em geral. As demonstrações destas propriedades são análogas às do caso unidimensional e por tal facto omiti-las-emos.

11.6. Cálculo de um integral duplo por integração unidimensional repetida

Na teoria da integração unidimensional, o segundo teorema fundamental do cálculo proporciona um método prático para o cálculo de integrais. O teorema que apresentamos a seguir conduz-nos a um resultado semelhante para a teoria da integração bidimensional; permite-nos calcular certos integrais duplos por intermédio de duas integrações unidimensionais sucessivas. O resultado é uma generalização da fórmula (11.3), a qual já foi demonstrada para funções em escada.

TEOREMA 11.5. *Seja f definida e limitada num rectângulo $Q = [a, b] \times [c, d]$ e admitase que f é integrável em Q . Para cada y fixo em $[c, d]$ admita-se que o integral unidimensional $\int_a^b f(x, y) dx$ existe e represente-se o seu valor por $A(y)$. Se o integral $\int_c^d A(y) dy$ existir, então será igual ao integral duplo $\iint_Q f$. Por outras palavras, tem-se a fórmula*

$$(11.6) \quad \iint_Q f(x, y) dx dy = \int_c^d \left[\int_a^b f(x, y) dx \right] dy.$$

Demonstração. Escolhamos duas quaisquer funções em escada s e t satisfazendo a $s \leq f \leq t$ em Q . Integramos relativamente a x e no intervalo $[a, b]$ temos

$$\int_a^b s(x, y) dx \leq A(y) \leq \int_a^b t(x, y) dx.$$

Visto que o integral $\int_c^d A(y) dy$ existe, podemos integrar ambas as desigualdades relativamente a y no intervalo $[c, d]$ e utilizar a Equação (11.3) para obter

$$\iint_Q s \leq \int_c^d A(y) dy \leq \iint_Q t.$$

relacionados com estes. Por fim indicamos como podem estes conceitos ser generalizados para um espaço a n dimensões.

11.2. Partições de rectângulos. Funções em escada

Seja Q um rectângulo, produto cartesiano dos dois intervalos fechados $[a, b]$ e $[c, d]$,

$$Q = [a, b] \times [c, d] = \{(x, y) \mid x \in [a, b] \text{ e } y \in [c, d]\}.$$

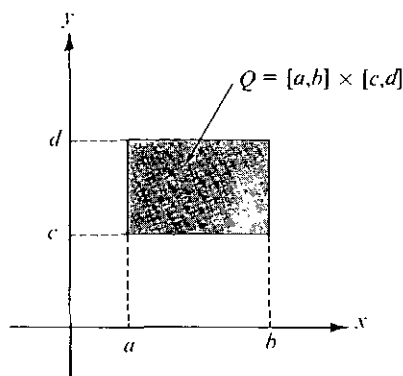


FIG. 11.1. Um rectângulo Q , produto cartesiano de dois intervalos.

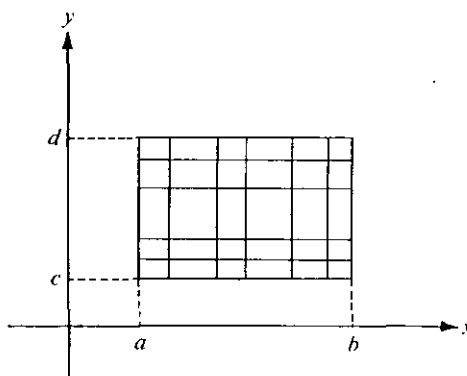


FIG. 11.2. Uma partição de um rectângulo Q .

Na figura 11.1 está representado um exemplo. Consideremos duas partições P_1 e P_2 , respectivamente de $[a, b]$ e $[c, d]$, sejam

$$P_1 = \{x_0, x_1, \dots, x_{n-1}, x_n\} \quad \text{e} \quad P_2 = \{y_0, y_1, \dots, y_{m-1}, y_m\},$$

onde $x_0 = a$, $x_n = b$, $y_0 = c$, $y_m = d$. O produto cartesiano $P_1 \times P_2$ define uma partição de Q . Visto P_1 decompôr $[a, b]$ em n subintervalos e P_2 decompôr $[c, d]$ em m subintervalos, a partição $P = P_1 \times P_2$ divide Q em mn rectângulos. A figura 11.2 mostra um exemplo de uma partição de Q em 30 subrectângulos. Uma partição P' de Q diz-se ser *mais fina* que P se $P \subseteq P'$, isto é, se cada ponto de P pertence também a P' .

O produto cartesiano de dois subintervalos abertos de P_1 e P_2 é um subrectângulo aberto (sem lados). Chama-se um subrectângulo aberto de P ou de Q .

DEFINIÇÃO DE FUNÇÃO EM ESCADA. Uma função f definida num rectângulo Q diz-se ser uma função em escada se existe uma partição P de Q tal que f seja constante em cada um dos subrectângulos abertos de P .

O gráfico de um exemplo de tal tipo de função está representado na figura 11.3. O gráfico é formado por placas rectangulares paralelas ao plano XOY . Uma função

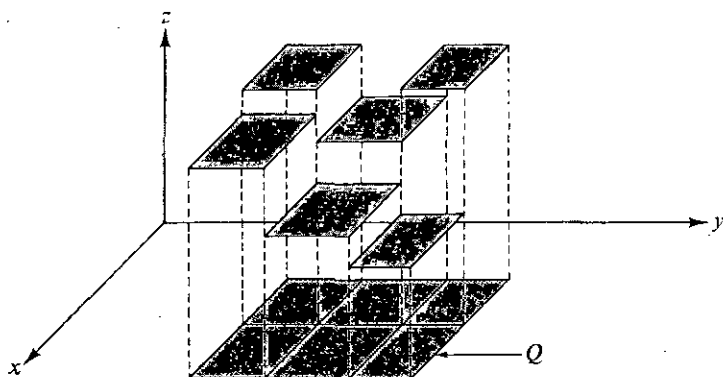


FIG. 11.3. Gráfica de uma função em escada definida num rectângulo

em escada também possui valores bem definidos em cada um dos pontos dos contornos dos subrectângulos, mas os valores nesses pontos não são relevantes para a teoria da integração.

Se f e g são duas funções em escada definidas num dado rectângulo Q , então a combinação linear $c_1 f + c_2 g$ é também uma função em escada. Com efeito, se P e P' são partições de Q tais que f seja constante nos subrectângulos abertos de P e g seja constante nos subrectângulos de P' , então $c_1 f + c_2 g$ é constante nos subrectângulos abertos da união $P \cup P'$ (que podemos chamar refinamento de P e P'). Assim, o conjunto de funções em escada definidas em Q constitui um espaço linear.

11.3. O integral duplo de uma função em escada

Sejam $P = P_1 \times P_2$ uma partição de um rectângulo Q em mn subrectângulos e f uma função em escada que é constante nos subrectângulos abertos de Q . Designemos por Q_{ij} o subrectângulo determinado por $[x_{i-1}, x_i]$ e $[y_{j-1}, y_j]$ e seja c_{ij} o valor constante de f no interior de Q_{ij} . Se f é positiva, o volume da caixa com base Q_{ij} e altura c_{ij} é o produto

$$c_{ij} \cdot (x_i - x_{i-1})(y_j - y_{j-1}).$$

Para qualquer função em escada f , positiva ou não, a soma de todos estes produtos torna-se como definição do integral duplo de f sobre Q . Assim, temos a seguinte definição:

DEFINIÇÃO DO INTEGRAL DUPLO DE UMA FUNÇÃO EM ESCADA. *Seja f uma função em escada que torna o valor constante c_{ij} no subrectângulo aberto $(x_{i-1}, x_i) \times (y_{j-1}, y_j)$ de um rectângulo Q . O integral duplo de f em Q define-se pela fórmula*

$$(11.1) \quad \iint_Q f = \sum_{i=1}^n \sum_{j=1}^m c_{ij} \cdot (x_i - x_{i-1})(y_j - y_{j-1}).$$

Portanto $\int_c^d A(y) dy$ é um número compreendido entre $\iint_Q s$ e $\iint_Q t$ para todas as funções em escada s e t aproximando f por baixo e por cima, respectivamente. Visto que f é integrável em Q , o único número com esta propriedade é o integral duplo de f em Q . Por conseguinte $\int_c^d A(y) dy = \iint_Q f$, o que demonstra a equação (11.6).

A fórmula (11.6) diz-se proporcionar um cálculo do integral duplo por integração repetida ou iterada. O processo consiste pois em integrar f com respeito a x de a a b (mantendo y fixo) e em segundo integrar o resultado com respeito a y de c a d . Se mudarmos a ordem de integração, temos uma fórmula análoga, a saber

$$(11.7) \quad \iint_Q f(x, y) dx dy = \int_a^b \left[\int_c^d f(x, y) dy \right] dx,$$

a qual é válida se supomos que $\int_c^d f(x, y) dy$ existe para cada x fixo em $[a, b]$ e é integrável em $[a, b]$.

11.7. Interpretação geométrica do integral duplo como um volume

O teorema 11.5 admite uma interpretação geométrica simples, representada na figura 11.4. Se f é não negativa, o conjunto S dos pontos (x, y, z) do 3-espaco com

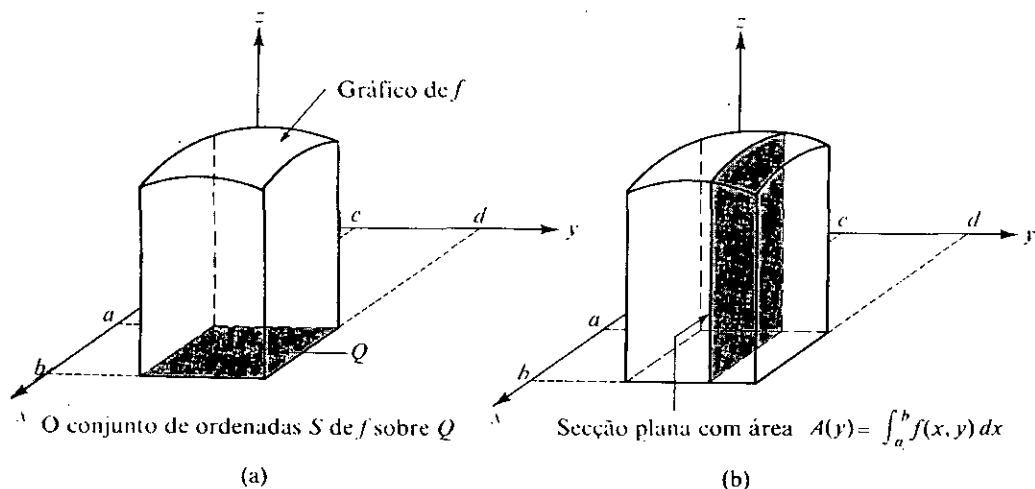


FIG. 11.4. O volume de S é o integral da área da secção plana:

$$v(S) = \int_c^d A(y) dy.$$

(x, y) em Q e $0 \leq z \leq f(x, y)$ chama-se o *conjunto de f sobre Q* . É formado por todos os pontos entre o rectângulo Q e a superfície $z = f(x, y)$. (Ver figura 11.4(a)). Para cada y

do intervalo $[c, d]$ o integral

$$A(y) = \int_a^b f(x, y) dx$$

é a área da secção plana de S definida por um plano paralelo ao plano XOZ (região sombreada na figura 11.4(b)). Visto que a área da secção $A(y)$ é integrável em $[c, d]$, o teorema 2.7 do Volume I diz-nos que o integral $\int_c^d A(y) dy$ é igual a $v(S)$, o volume de S . Assim, para integrandos não negativos, o teorema 11.5 mostra que o volume de conjunto de ordenadas de f sobre Q é igual ao integral duplo $\iint_Q f$.

A equação (11.7) dá-nos outra maneira de calcular o volume do conjunto de ordenadas. Aqui integramos a área das secções planas produzidas por planos paralelos ao plano YOZ .

11.8. Exemplos resolvidos

Vamos de seguida efectuar aplicações do teorema 11.5 por meio de dois exemplos numéricos.

EXEMPLO 1. Se $Q = [-1, 1] \times [0, \pi/2]$, calcular $\iint_Q (x \operatorname{sen} y - ye^x) dx dy$, admitindo que o integral existe. A região de integração está representada na figura 11.5.

Resolução. Integramos em primeiro lugar em relação a x e designando o resultado por $A(y)$, temos

$$A(y) = \int_{-1}^1 (x \operatorname{sen} y - ye^x) dx = \left(\frac{x^2}{2} \operatorname{sen} y - ye^x \right) \Big|_{x=-1}^{x=1} = -ey + y/e.$$

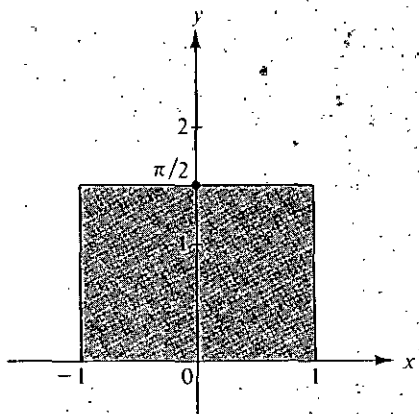


FIG. 11.5. A região de integração para o Exemplo 1.

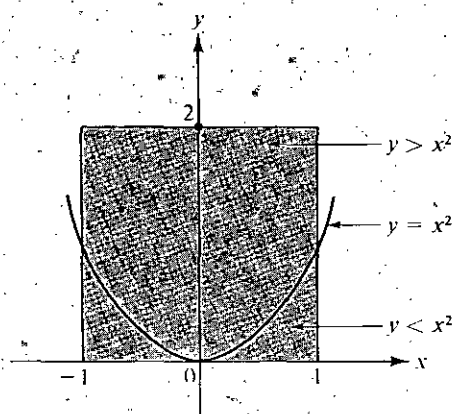


FIG. 11.6. A região de integração para o Exemplo 2.

Aplicando o teorema 11.5 encontramos

$$\begin{aligned}\iint_Q (x \operatorname{sen} y - ye^x) dx dy &= \int_0^{\pi/2} A(y) dy = \int_0^{\pi/2} (-ey + y/e) dy \\ &= (1/e - e) \int_0^{\pi/2} y dy = (1/e - e)\pi^2/8.\end{aligned}$$

Como prova dos cálculos podemos integrar em primeiro lugar em relação a y ;

$$\begin{aligned}\iint_Q (x \operatorname{sen} y - ye^x) dx dy &= \int_{-1}^1 \left[\int_0^{\pi/2} (x \operatorname{sen} y - ye^x) dy \right] dx \\ &= \int_{-1}^1 \left(-x \cos y - \frac{1}{2} y^2 e^x \right) \Big|_{y=0}^{y=\pi/2} dx \\ &= \int_{-1}^1 (-\pi^2 e^x/8 + x) dx = (1/e - e)\pi^2/8.\end{aligned}$$

EXEMPLO 2. Se $Q = [-1, 1] \times [0, 2]$, calcular o integral duplo $\iint_Q \sqrt{|y - x^2|} dx dy$, admitido a sua existência.

Resolução. Se integramos em primeiro lugar em relação a y e designamos o resultado por $H(x)$, temos $H(x) = \int_0^2 \sqrt{|y - x^2|} dy$. A região de integração é o rectângulo representado na figura 11.6. A parábola $y = x^2$ está também desenhada devido à presença de $|y - x^2|$ no integrando. Acima desta parábola temos $y > x^2$ e abaixo $y < x^2$. Isto sugere que devidamos o integral de $H(x)$ do modo seguinte:

$$H(x) = \int_0^2 \sqrt{|y - x^2|} dy = \int_0^{x^2} \sqrt{x^2 - y} dy + \int_{x^2}^2 \sqrt{y - x^2} dy.$$

Lembramos que x é considerado como constante em cada um destes integrais. No primeiro integral efectuamos a mudança de variável $t = x^2 - y$ e no segundo fazemos $t = y - x^2$. Isto dá-nos

$$H(x) = \int_0^2 \sqrt{|y - x^2|} dy = -\int_{x^2}^0 \sqrt{t} dt + \int_0^{2-x^2} \sqrt{t} dt = \frac{2}{3} x^3 + \frac{2}{3} (2 - x^2)^{3/2}.$$

Aplicando o teorema 11.5 temos

$$\begin{aligned}\iint_Q \sqrt{|y - x^2|} dx dy &= \int_{-1}^1 \left[\frac{2}{3} x^3 + \frac{2}{3} (2 - x^2)^{3/2} \right] dx = \frac{4}{3} \int_0^1 (2 - x^2)^{3/2} dx \\ &= \frac{1}{3} \left[x(2 - x^2)^{3/2} + 3x\sqrt{2 - x^2} + 6 \arcsen\left(\frac{x}{\sqrt{2}}\right) \right] \Big|_0^1 = \frac{4}{3} + \frac{\pi}{2}.\end{aligned}$$

O mesmo resultado se obteria integrando em primeiro lugar em relação a x , mas os cálculos são mais complicados.

11.9. Exercícios

Calcular os integrais duplos por integração repetida, admitindo a existência de cada integral

1. $\iint_Q xy(x+y) dx dy$, onde $Q = [0, 1] \times [0, 1]$.

2. $\iint_Q (x^3 + 3x^2y + y^3) dx dy$, onde $Q = [0, 1] \times [0, 1]$.

3. $\iint_Q (\sqrt{y} + x - 3xy^2) dx dy$, onde $Q = [0, 1] \times [1, 3]$.

4. $\iint_Q \sin^2 x \sin^2 y dx dy$, onde $Q = [0, \pi] \times [0, \pi]$.

5. $\iint_Q \sin(x+y) dx dy$, onde $Q = [0, \pi/2] \times [0, \pi/2]$.

6. $\iint_Q |\cos(x+y)| dx dy$, onde $Q = [0, \pi] \times [0, \pi]$.

7. $\iint_Q f(x+y) dx dy$, onde $Q = [0, 2] \times [0, 2]$, e $f(t)$ representa o maior inteiro $\leq t$.

8. $\iint_Q y^{-3} e^{tx/y} dx dy$, onde $Q = [0, t] \times [1, t]$, $t > 0$.

9. Se Q é um rectângulo, mostrar que um integral duplo da forma $\iint_Q f(y)g(y) dx dy$ é igual ao produto de dois integrais unidimensionais. Estabelecer as hipóteses utilizadas respeitantes à existência.

10. Seja f definida num rectângulo $Q = [0, 1] \times [0, 1]$ do modo seguinte:

$$f(x, y) = \begin{cases} 1 - x - y & \text{se } x + y \leq 1 \\ 0 & \text{nas restantes hipóteses.} \end{cases}$$

Representar o conjunto de ordenadas de f sobre Q e calcular o volume deste conjunto de ordenadas por integração dupla. (Admitir a existência do integral.)

11. Resolver o Exercício 10 quando

$$f(x, y) = \begin{cases} x + y & \text{se } x^2 \leq y \leq 2x^2 \\ 0 & \text{nas restantes hipóteses.} \end{cases}$$

12. Resolver o Exercício 10 quando $Q = [-1, 1] \times [-1, 1]$ e

$$f(x, y) = \begin{cases} x^2 + y^2 & \text{se } x^2 + y^2 \leq 1, \\ 0 & \text{nas restantes hipóteses.} \end{cases}$$

13. Seja f definida num rectângulo $Q = [1, 2] \times [1, 4]$ do modo seguinte:

$$f(x, y) = \begin{cases} (x + y)^{-2} & \text{se } x \leq y \leq 2x, \\ 0 & \text{nas restantes hipóteses.} \end{cases}$$

Representar a parte de Q na qual f é não nula e calcular o valor do integral duplo $\iint_Q f$, suposto existir

14. Seja f definida no rectângulo $Q = [0, 1] \times [0, 1]$ do modo seguinte:

$$f(x, y) = \begin{cases} 1 & \text{se } x = y, \\ 0 & \text{se } x \neq y. \end{cases}$$

Provar que o integral duplo $\iint_Q f$ existe e é igual a 0.

11.10. Integrabilidade de funções contínuas

O teorema da continuidade uniforme (teorema 9.10) pode utilizarse para provar a integrabilidade de uma função que seja contínua num rectângulo.

TEOREMA 11.6. INTEGRABILIDADE DE FUNÇÕES CONTÍNUAS. *Se uma função f é contínua num rectângulo $Q = [a, b] \times [c, d]$, então f é integrável em Q . Além disso, o valor do integral pode obter-se por integração unidimensional repetida*

$$(11.8) \quad \iint_Q f = \int_c^d \left[\int_a^b f(x, y) dx \right] dy = \int_a^b \left[\int_c^d f(x, y) dy \right] dx.$$

Demonstração. O teorema (9.8) mostra que f é limitada em Q , pelo que f admite um integral superior e um integral inferior. Pretendemos provar que $I(f) = \bar{I}(f)$. Escolhamos $\epsilon > 0$. Pelo teorema da continuidade uniforme, para esta escolha de ϵ existe uma partição P de Q num número finito (digamos n) de subrectângulos Q_1, \dots, Q_n tais que a oscilação de f em cada subrectângulo seja menor que ϵ . Representemos por $M_k(f)$ e $m_k(f)$, respectivamente, os valores máximo e mínimo absolutos de f em Q_k . Então temos

$$M_k(f) - m_k(f) < \epsilon$$

para cada $k = 1, 2, \dots, n$. Sejam agora s e t duas funções em escada definidas no interior de cada Q_k do modo seguinte:

$$s(x) = m_k(f), \quad t(x) = M_k(f) \quad \text{se } x \in \text{int } Q_k.$$

Nos pontos fronteira definimos $s(x) = m$ e $t(x) = M$, em que m e M são, respectivamente, o mínimo e máximo valores absolutos de f em Q . Então tem-se $s \leq f \leq t$ para todo x em Q . Do mesmo modo temos

$$\iint_Q s = \sum_{k=1}^n m_k(f) a(Q_k) \quad \text{e} \quad \iint_Q t = \sum_{k=1}^n M_k(f) a(Q_k),$$

em que $a(Q_k)$ é a área do rectângulo Q_k . A diferença destes dois integrais é

$$\iint_Q t - \iint_Q s = \sum_{k=1}^n \{M_k(f) - m_k(f)\} a(Q_k) < \epsilon \sum_{k=1}^n a(Q_k) = \epsilon a(Q),$$

com $a(Q)$ a área de Q . Uma vez que $\int_Q s \leq I(f) \leq \overline{I}(f) \leq \int_Q t$, resulta a desigualdade

$$0 \leq \overline{I}(f) - I(f) \leq \epsilon a(Q).$$

Fazendo $\epsilon \rightarrow 0$ vemos que $\overline{I}(f) = I(f)$, pelo que f é integrável em Q .

Provamos a seguir que o integral duplo é igual aos dois integrais simples sucessivos (integrais repetidos) de (11.8). Para cada y fixo em $[c, d]$ o integral simples $\int_a^b f(x, y) dx$ existe, uma vez que o integrando é contínuo em Q . Seja $A(y) = \int_a^b f(x, y) dx$. Interessa provar que A é contínua em $[c, d]$. Se y e y_1 são dois pontos quaisquer de $[c, d]$ temos

$$A(y) - A(y_1) = \int_a^b \{f(x, y) - f(x, y_1)\} dx$$

onde resulta

$$|A(y) - A(y_1)| \leq (b - a) \max_{a \leq x \leq b} |f(x, y) - f(x, y_1)| = (b - a) |f(x_1, y) - f(x_1, y_1)|$$

onde x_1 é um ponto de $[a, b]$ no qual $|f(x, y) - f(x, y_1)|$ atinge o seu valor máximo. Esta desigualdade mostra que $A(y) \rightarrow A(y_1)$ quando $y \rightarrow y_1$, pelo que A é contínua em y_1 . Deste modo o integral $\int_c^d A(y) dy$ existe e, pelo teorema 11.5, é igual a $\int_Q f$. Um raciocínio semelhante é ainda válido quando se inverte a ordem de integração.

11.11. Integrabilidade de funções limitadas com descontinuidades

Seja f definida e limitada num rectângulo Q . No teorema 11.6 provámos que o integral duplo de f sobre Q existe se f for contínua em todos os pontos de Q . Nesta secção vamos demonstrar que o integral também existe se f tem descontinuidades em Q , desde que o conjunto das descontinuidades não seja demasiado grande. Para medir a grandeza do conjunto de descontinuidades introduzimos o seguinte conceito.

DEFINIÇÃO DE UM CONJUNTO LIMITADO DE MEDIDA NULA. *Seja A um subconjunto limitado do plano. Diz-se que o conjunto A tem medida nula se para todo $\epsilon > 0$ existe um conjunto finito de rectângulos cuja união contém A e cuja soma das áreas não excede ϵ .*

Por outras palavras, um conjunto plano limitado de medida nula pode ser *contido* por uma reunião de rectângulos cuja área total seja arbitrariamente pequena.

As seguintes proposições relativas a conjuntos limitados de medida nula são consequências imediatas desta definição. As demonstrações são, por esse motivo, deixadas ao leitor como exercícios.

- (a) Qualquer conjunto finito de pontos no plano tem medida nula.
- (b) A reunião de um número finito de conjuntos limitados de medida nula tem também medida nula
- (c) Todo o subconjunto de um conjunto de medida nula tem medida nula.
- (d) Todo o segmento de recta tem medida nula.

TEOREMA 11.7. *Seja f definida e limitada num rectângulo $Q = [a, b] \times [c, d]$. Se o conjunto das descontinuidades de f em Q é um conjunto de medida nula, então o integral duplo $\iint_Q f$ existe.*

Demonstração. Seja $M > 0$, tal que $|f| \leq M$ em Q . Represente D o conjunto das descontinuidades de f em Q . Dado $\delta > 0$, seja P uma partição de Q tal que a soma das áreas de todos os subrectângulos de P , que contêm pontos de D , seja menor que δ . (Isto é normal porque D tem medida nula). Definamos nestes subrectângulos funções em escada s e t do modo seguinte:

$$s(x) = -M, \quad t(x) = M.$$

Nos restantes subrectângulos de P definamos s e t tal como na demonstração do teorema 11.6. Será então $s \leq f \leq t$ em todo Q . Argumentando do mesmo modo que na demonstração do teorema 11.6 obtemos a desigualdade

$$(11.9) \quad \iint_Q t - \iint_Q s \leq \epsilon a(Q) + 2M\delta.$$

O primeiro termo, $\epsilon a(Q)$, resulta do cálculo do integral de $t - s$ estendido aos rectângulos contendo unicamente pontos de continuidade de f ; o segundo termo, $2M\delta$, resulta do cálculo do integral de $t - s$ estendido aos rectângulos que contêm pontos de D . De (11.9) obtemos a desigualdade

$$0 \leq I(f) - J(f) \leq \epsilon a(Q) + 2M\delta.$$

Fazendo $\epsilon \rightarrow 0$ obtemos $0 \leq \bar{I}(f) - J(f) \leq 2M\delta$. Uma vez que δ é arbitrário isto implica que $J(f) = \bar{I}(f)$, pelo que f é integrável em Q .

11.12. Integrais duplos estendidos a regiões mais gerais

Até aqui os integrais duplos foram definidos unicamente para domínios de integração rectangulares. Não é contudo difícil estender o conceito a domínios mais gerais.

Seja S uma região limitada, e encerremo-la num rectângulo Q . Seja f definida e limitada em S . Definamos uma nova função \tilde{f} em Q do modo seguinte:

$$(11.10) \quad \tilde{f}(x, y) = \begin{cases} f(x, y) & \text{se } (x, y) \in S, \\ 0 & \text{se } (x, y) \in Q - S. \end{cases}$$

Por outras palavras, estendemos a definição de f a todo o rectângulo Q fazendo com que f seja nula no exterior de S . Pergunta-se agora se sim ou não a função \tilde{f} é integrável em Q . Caso afirmativo, dizemos que f é integrável em S e que, *por definição*,

$$\iint_S f = \iint_Q \tilde{f}.$$

Consideremos em primeiro lugar conjuntos S de pontos no plano XOY definidos do modo seguinte:

$$S = \{(x, y) \mid a \leq x \leq b \text{ e } \varphi_1(x) \leq y \leq \varphi_2(x)\},$$

onde φ_1 e φ_2 são funções contínuas num intervalo fechado $[a, b]$, e satisfazendo a $\varphi_1 \leq \varphi_2$. Um exemplo de uma tal região, a qual pode designar-se do Tipo I, está representado na figura 11.7. Numa região do Tipo I, para cada ponto t em $[a, b]$, a recta $x = t$ intersecta S segundo um segmento que une a curva $y = \varphi_1(x)$ a $y = \varphi_2(x)$. Uma tal região é limitada, porque φ_1 e φ_2 são contínuas e por conseguinte limitadas em $[a, b]$.

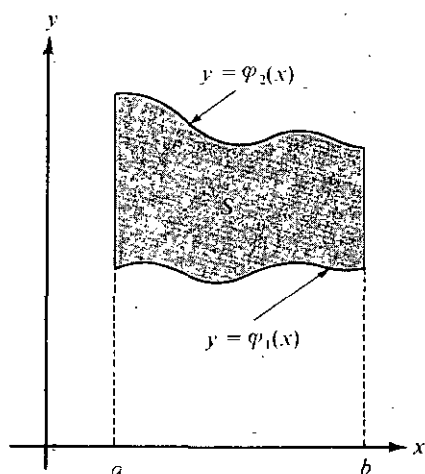


FIG. 11.7 Uma região S do tipo I

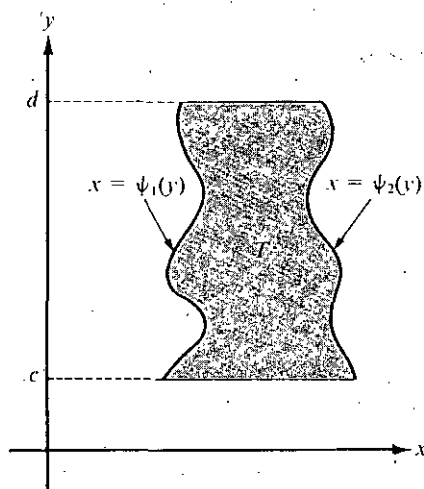


FIG. 11.8 Uma região T do tipo II.

Outro tipo de região T (tipo II) pode ser definido do modo seguinte:

$$T = \{(x, y) \mid c \leq y \leq d \quad \text{e} \quad \psi_1(y) \leq x \leq \psi_2(y)\},$$

onde ψ_1 e ψ_2 são contínuas num intervalo $[c, d]$ sendo $\psi_1 \leq \psi_2$. Como exemplo temos a figura 11.8. Neste caso rectas paralelas a OX intersectam T segundo segmentos que unem as curvas $x_1 = \psi_1(y)$ e $x_2 = \psi_2(y)$. Regiões do tipo II são também limitadas. Todas as regiões que se considerem, serão de um destes dois tipos ou poderão ser decompostas num número finito de partes, cada uma das quais será de um destes dois tipos.

Seja f definida e limitada numa região S do tipo I. Encerremos S num rectângulo Q e definamos \tilde{f} em Q como está indicado na equação (11.10). As descontinuidades de \tilde{f} em Q serão as de f em S , mais aqueles pontos da fronteira de S nos quais f não é nula. A fronteira de S é constituída pelos gráficos de φ_1 , de φ_2 e pelos dois segmentos de recta, paralelos a OY , que unem as extremidades dos gráficos. Cada um desses segmentos de recta tem medida nula. O teorema que apresentamos a seguir prova que cada um dos gráficos tem também medida nula.

TEOREMA 11.8. *Se φ é uma função real contínua num intervalo $[a, b]$, então o gráfico de φ tem medida nula.*

Demonstração. Seja A o gráfico de φ , isto é,

$$A = \{(x, y) \mid y = \varphi(x) \quad \text{e} \quad a \leq x \leq b\}.$$

Escolhamos $\epsilon > 0$, e apliquemos o teorema da continuidade uniforme (teorema 3.13 do Volume I) para obtermos uma partição P de $[a, b]$ num número finito de subintervalos tais que a oscilação de φ em cada subintervalo de P seja menor que $\epsilon/(b-a)$. Portanto, relativamente a cada subintervalo de P o gráfico de φ está situado num rectângulo de altura $\epsilon/(b-a)$. Logo todo o gráfico de φ está contido numa união de um

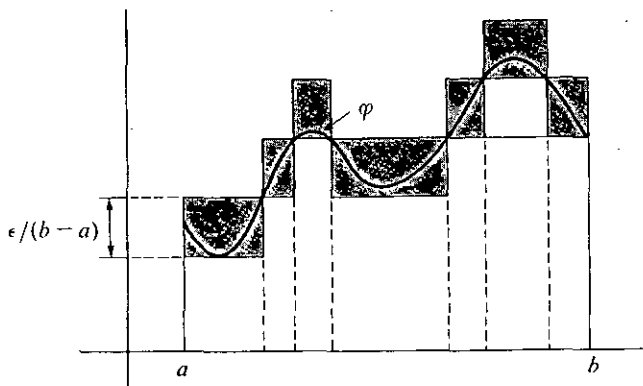


FIG. 11.9. Demonstração de que o gráfico de uma função contínua φ tem medida nula.

número finito de rectângulos, cuja soma das áreas é ϵ . (Ver figura 11.9.) Isto prova que o gráfico de φ tem medida nula.

O teorema que apresentamos a seguir prova que o integral duplo $\iint_S f$ existe se f for contínua em $\text{int } S$ (o interior de S). Este é o conjunto

$$\text{int } S = \{(x, y) \mid a < x < b \quad \text{e} \quad \varphi_1(x) < y < \varphi_2(x)\}.$$

TEOREMA 11.9. *Seja S uma região do Tipo I, compreendida entre os gráficos de φ_1 e φ_2 . Admita-se que f está definida e limitada em S e que f é contínua no interior de S . Então o integral duplo $\iint_S f$ existe e pode calcular-se mediante integração unidimensional repetida,*

$$(11.11) \quad \iint_S f(x, y) \, dx \, dy = \int_a^b \left[\int_{\varphi_1(x)}^{\varphi_2(x)} f(x, y) \, dy \right] dx.$$

Demonstração. Sejam $Q = [a, b] \times [c, d]$ um rectângulo que contém S , e \tilde{f} uma função definida por (11.10). Os únicos pontos de descontinuidade possíveis para \tilde{f} são os pontos da fronteira de S . Uma vez que a fronteira de S tem medida nula, \tilde{f} é integrável em Q . Para cada x fixo em (a, b) o integral $\int_c^d \tilde{f}(x, y) \, dy$ existe, uma vez que o integrando tem quando muito duas descontinuidades em $[c, d]$. Aplicando a versão do teorema 11.5 dada pela equação (11.7) temos

$$(11.12) \quad \iint_Q \tilde{f} = \int_a^b \left[\int_c^d \tilde{f}(x, y) \, dy \right] dx.$$

Se $\varphi_1(x) \leq y \leq \varphi_2(x)$ então $\tilde{f}(x, y) = f(x, y)$, enquanto que $\tilde{f}(x, y) = 0$ para os restantes valores de y em $[c, d]$. Portanto

$$\int_c^d \tilde{f}(x, y) \, dy = \int_{\varphi_1(x)}^{\varphi_2(x)} f(x, y) \, dy$$

pelo que a equação (11.12) implica (11.11).

Existe, naturalmente, um teorema análogo para uma região T do tipo II. Se f é definida e limitada em T e contínua no interior de T , então f é integrável em T e a fórmula para a integração repetida vem

$$(11.13) \quad \iint_T f(x, y) \, dx \, dy = \int_c^d \left[\int_{\varphi_1(y)}^{\varphi_2(y)} f(x, y) \, dx \right] dy.$$

Algumas regiões são simultaneamente do tipo I e do tipo II. (Regiões limitadas por circunferências e elipses são disso exemplo.) Neste caso a ordem de integração é irrelevante e podemos escrever

$$\int_a^b \left[\int_{\varphi_1(x)}^{\varphi_2(x)} f(x, y) dy \right] dx = \int_c^d \left[\int_{\varphi_1(y)}^{\varphi_2(y)} f(x, y) dx \right] dy.$$

Em alguns casos um desses integrais pode ser mais fácil de calcular do que o outro; torna-se por isso vantajoso examiná-los antes de proceder ao cálculo efectivo de um integral duplo.

11.13. Aplicações a áreas e volumes

Seja S uma região do Tipo I definida por

$$S = \{(x, y) \mid a \leq x \leq b \quad \text{e} \quad \varphi_1(x) \leq y \leq \varphi_2(x)\}.$$

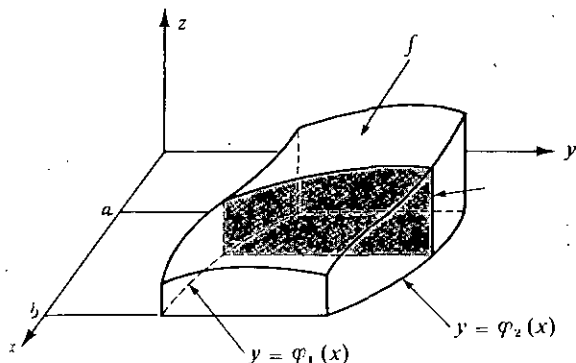
Aplicando o teorema 11.9, com $f(x, y) = 1$ para todo (x, y) em S , obtemos

$$\iint_S dx dy = \int_a^b [\varphi_2(x) - \varphi_1(x)] dx.$$

Pelo teorema 2.1 do Volume I, o integral do segundo membro é igual à área da região S . Assim concluímos que os integrais duplos podem ser utilizados no cálculo de áreas.

Se f é não negativa, o conjunto de pontos (x, y, z) no espaço tridimensional tais que $(x, y) \in S$ e $0 \leq z \leq f(x, y)$ diz-se o *conjunto de ordenadas de f sobre S* . Na figura 11.10 está representado um exemplo. Se f é não e contínua em S , o integral

$$\int_{\varphi_1(x)}^{\varphi_2(x)} f(x, y) dy$$



$$\text{plana } \int_{\varphi_1(x)}^{\varphi_2(x)} f(x, y) dy$$

FIG. 11.10. O integral $\int_{\varphi_1(x)}^{\varphi_2(x)} f(x, y) dy$ é a área de uma secção plana do conjunto de ordenadas.

O integral $\int_a^b \left[\int_{\varphi_1(x)}^{\varphi_2(x)} f(x, y) dy \right] dx$ é o volume do conjunto de ordenadas.

representa a área da secção plana do referido conjunto, definida por um plano paralelo ao plano YOZ (a parte sombreada da figura 11.10). A fórmula (11.11) do teorema 11.9 mostra que o integral duplo de f sobre S é igual ao integral da área das secções. Por conseguinte, o integral duplo $\iint_S f$ é igual ao volume do conjunto de ordenadas de f sobre S . (Ver teorema 2.7 do Volume I, p. 113.)

Mais geralmente, se f e g são ambas contínuas em S com $f \leq g$, então o integral duplo $\iint_S (g - f)$ é igual ao volume do sólido compreendido entre os gráficos das funções f e g . É evidente que observações análogas se aplicam a regiões do tipo II.

11.14. Exemplos resolvidos

EXEMPLO 1. Calcular o volume do sólido constituído pelo elipsoide

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1.$$

Resolução. O sólido em questão está compreendido entre os gráficos das funções f e g , sendo

$$g(x, y) = c\sqrt{1 - x^2/a^2 - y^2/b^2} \quad \text{e} \quad f(x, y) = -g(x, y).$$

Aqui $(x, y) \in S$, sendo S a região elíptica dada por

$$S = \left\{ (x, y) \left| \frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1 \right. \right\}.$$

Aplicando o teorema 11.9, e tendo em conta a simetria, vem o volume do sólido elipsoidal dado por

$$V = \iint_S (g - f) = 2 \iint_S g = 8c \int_0^a \left[\int_0^{b\sqrt{1-x^2/a^2}} \sqrt{1 - x^2/a^2 - y^2/b^2} dy \right] dx.$$

Seja $A = \sqrt{1 - x^2/a^2}$. Então o integral do parêntesis escreve-se

$$\int_0^{bA} \sqrt{A^2 - y^2/b^2} dy = A \int_0^{bA} \sqrt{1 - y^2/(Ab)^2} dy.$$

Efectuando a mudança de variável $y = Ab \sin t$, $dy = Ab \cos t dt$, encontramos para o último integral

$$A^2 b \int_0^{\pi/2} \cos^2 t dt = \frac{\pi}{4} A^2 b = \frac{\pi b}{4} \left(1 - \frac{x^2}{a^2} \right).$$

Portanto

$$V = 8c \int_0^a \frac{\pi b}{4} \left(1 - \frac{x^2}{a^2}\right) dx = \frac{4}{3} \pi abc.$$

No caso particular em que $a = b = c$ o sólido é uma esfera de raio a cujo volume é $\frac{4}{3} \pi a^3$.

EXEMPLO 2. O integral duplo de uma função positiva f , $\iint_S f(x, y) dx dy$, reduz-se ao cálculo dos integrais repetidos

$$\int_0^1 \left[\int_{x^2}^x f(x, y) dy \right] dx.$$

Determinar a região S e permutar a ordem de integração.

Resolução. Para cada x fixo entre 0 e 1, a integração com respeito a y efectua-se no intervalo de x^2 a x . Isto significa que a região é do Tipo I e está limitada pelas duas curvas $y = x^2$ e $y = x$. A região S é o conjunto de pontos situados entre duas curvas

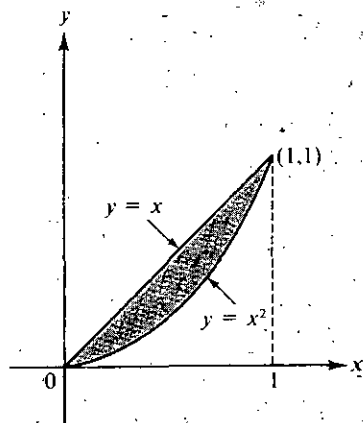


FIG. 11.11. Exemplo 2

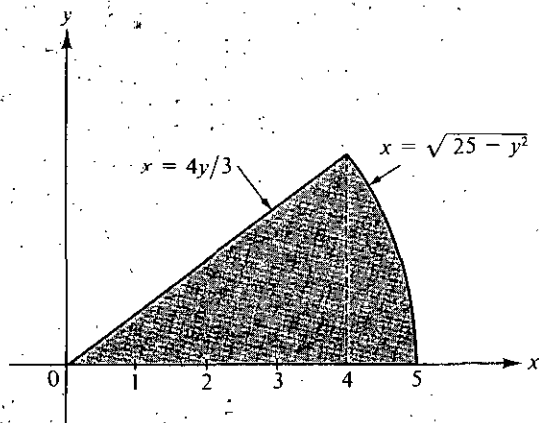


FIG. 11.12. Exemplo 3.

e sobre o intervalo $[0, 1]$. (Ver figura 11.11). Uma vez que S é também do tipo II podemos inverter a ordem de integração obtendo

$$\int_0^1 \left[\int_y^{\sqrt{y}} f(x, y) dx \right] dy.$$

EXEMPLO 3. Um integral duplo de uma função positiva f , $\iint_S f(x, y) dx dy$, reduz-se ao cálculo dos integrais repetidos:

$$\int_0^3 \left[\int_{4y/3}^{\sqrt{25-y^2}} f(x, y) dx \right] dy.$$

Determinar a região S e permutar a ordem de integração.

Resolução. Para cada y fixo entre 0 e 3, a integração com respeito a x é referida ao intervalo de $4y/3$ a $\sqrt{25-y^2}$. Portanto a região S é do tipo II e está situada entre as duas curvas $x = 4y/3$ e $x = \sqrt{25-y^2}$. Esta região, representada na figura 11.12, é um sector circular. Quando se inverte a ordem, a região deve dividir-se em duas, ambas do tipo I; o resultado é a soma de dois integrais.

$$\int_0^4 \left[\int_0^{3x/4} f(x, y) dy \right] dx + \int_4^5 \left[\int_0^{\sqrt{25-x^2}} f(x, y) dy \right] dx.$$

11.15. Exercícios

Nos Exercícios 1 a 5, esboçar um desenho da região de integração e calcular o integral duplo.

1. $\iint_S x \cos(x+y) dx dy$, sendo S a região triangular cujos vértices são $(0, 0)$, $(\pi, 0)$, (π, π) .

2. $\iint_S (1+x) \sin y dx dy$, sendo S o trapézio com vértices $(0, 0)$, $(1, 0)$, $(1, 2)$, $(0, 1)$.

3. $\iint_S e^{x+y} dx dy$, sendo $S = \{(x, y) \mid |x| + |y| \leq 1\}$.

4. $\iint_S x^2 y^2 dx dy$, sendo S a região do primeiro quadrante entre as hipérbolas $xy = 1$ e $xy = 2$ e as duas rectas $y = x$ e $y = 4x$.

5. $\iint_S (x^2 - y^2) dx dy$, sendo S a região limitada pela curva $y = \sin x$ e o intervalo $[0, \pi]$.

6. Uma pirâmide está limitada pelos três planos coordenados e o plano $x + 2y + 3z = 6$. Fazer um desenho do sólido referido e calcular o seu volume por integração dupla.

7. Um sólido está limitado pela superfície $z = x^2 - y^2$, o plano XOY e os planos $x = 1$ e $x = 3$. Fazer um desenho do referido sólido e calcular o seu volume por intermédio de um integral duplo.

8. Calcular, por um integral duplo, o volume do conjunto entre, f e S se:

(a) $f(x, y) = x^2 + y^2$ e $S = \{(x, y) \mid |x| \leq 1, |y| \leq 1\}$.

(b) $f(x, y) = 3x + y$ e $S = \{(x, y) \mid 4x^2 + 9y^2 \leq 36, x > 0, y > 0\}$.

(c) $f(x, y) = y + 2x + 20$ e $S = \{(x, y) \mid x^2 + y^2 \leq 16\}$.

Nos Exercícios 9 a 18 supomos que o integral duplo de uma função positiva f , estendido a uma região S , se reduz aos integrais repetidos dados. Em cada exercício traçar um esboço de S e permutar a ordem de integração.

9. $\int_0^1 \left[\int_0^y f(x, y) dx \right] dy.$

10. $\int_0^2 \left[\int_{y^2}^{2y} f(x, y) dx \right] dy.$

11. $\int_1^4 \left[\int_{\sqrt{x}}^2 f(x, y) dy \right] dx.$

$$12. \int_1^2 \left[\int_{2-x}^{\sqrt{2x-x^2}} f(x, y) dy \right] dx.$$

$$13. \int_{-6}^2 \left[\int_{(x^2-4)/4}^{2-x} f(x, y) dy \right] dx.$$

$$14. \int_1^e \left[\int_0^{\log x} f(x, y) dy \right] dx.$$

$$15. \int_{-1}^1 \left[\int_{-\sqrt{1-x^2}}^{1-x^2} f(x, y) dy \right] dx.$$

$$16. \int_0^1 \left[\int_{x^3}^{x^2} f(x, y) dy \right] dx.$$

$$17. \int_0^{\pi} \left[\int_{-\sin(x/2)}^{\sin x} f(x, y) dy \right] dx.$$

$$18. \int_0^4 \left[\int_{-\sqrt{4-y}}^{(y-4)/2} f(x, y) dx \right] dy.$$

19. Ao calcular-se, por dupla integração, o volume V do sólido situado abaixo do parabolóide $z = x^2 + y^2$ e acima de uma região S do plano XOY , obteve-se a seguinte soma de integrais repetidos

$$V = \int_0^1 \left[\int_0^y (x^2 + y^2) dx \right] dy + \int_1^2 \left[\int_0^{2-y} (x^2 + y^2) dx \right] dy.$$

Desenhar a região S e exprimir V por integrais repetidos nos quais a ordem de integração esteja invertida. Efectuar, também, a integração e calcular V .

20. Ao calcular-se, por dupla integração, o volume V do sólido situado sob a superfície $z = f(x, y)$ e acima de uma região S do plano XOY obteve-se a seguinte soma de integrais:

$$V = \int_0^{a \operatorname{sen} c} \left[\int_{\sqrt{a^2-y^2}}^{\sqrt{b^2-y^2}} f(x, y) dx \right] dy + \int_{a \operatorname{sen} c}^b \left[\int_{y \cot c}^{\sqrt{b^2-y^2}} f(x, y) dx \right] dy.$$

Dado que $0 < a < b$ e $0 < c < \pi/2$, desenhar S dando as equações de todas as curvas que formam a sua fronteira.

21. Quando por um integral duplo se calculou o volume V do sólido sob a superfície $z = f(x, y)$ e acima da região S de XOY , obteve-se a seguinte soma de integrais repetidos

$$V = \int_1^2 \left[\int_x^{x^3} f(x, y) dy \right] dx + \int_2^8 \left[\int_x^8 f(x, y) dy \right] dx.$$

(a) Desenhar a região S e exprimir V por integrais repetidos nos quais a ordem de integração seja invertida.

(b) Efectuar a integração e calcular V quando $f(x, y) = e^x(x/y)^{1/2}$.

22. Seja $A = \int_0^1 e^{-t} dt$ e $B = \int_{1/2}^1 e^{-t} dt$. Calcular o integral

$$I = 2 \int_{-1/2}^1 \left[\int_0^x e^{-v^2} dy \right] dx$$

Integrais múltiplos

em função de A e B . Existem inteiros e positivos m e n tais que

$$I = mA - nB + e^{-1} - e^{-1/4}.$$

Servir-se desta fórmula para testar a resposta.

23. Obtem-se um cone unindo todos os pontos de uma região plana S com um ponto não situado no plano de S . Designando por A a área de S e por h a altura do cone, provar que:
- (a) A área da secção produzida por um plano paralelo à base e à distância t do vértice é $(t/h)^2 A$, se $0 \leq t \leq h$.
- (b) O volume do cone é $\frac{1}{3} Ah$.
24. Inverter a ordem de integração para derivar a fórmula

$$\int_0^a \left[\int_0^y e^{m(a-x)} f(x) dx \right] dy = \int_0^a (a-x) e^{m(a-x)} f(x) dx,$$

onde a e m são constantes, $a > 0$.

11.16. Outras aplicações dos integrais duplos

Já vimos que os integrais duplos podem servir no cálculo de volumes de sólidos e áreas de regiões planas. Muitos outros conceitos tais como massa, centro de massa, e momento de inércia podem ser definidos e calculados com a ajuda de integrais duplos. Esta secção é dedicada a uma breve discussão destes tópicos, os quais são de grande importância na física e engenharia.

Designemos por P o vector que une a origem a um ponto arbitrário no espaço tridimensional. Se n massas positivas m_1, m_2, \dots, m_n estão localizadas nos pontos P_1, P_2, \dots, P_n , respectivamente, o *centro de massa* do sistema define-se como sendo o ponto C dado pela equação vectorial

$$C = \frac{\sum_{k=1}^n m_k P_k}{\sum_{k=1}^n m_k}.$$

o denominador, $\sum m_k$, chama-se a *massa total* do sistema.

Se cada massa m_k for transladada, segundo um vector A , para um novo ponto Q_k em que $Q_k = P_k + A$, o centro de massa é também transladado de A , visto verificar-se

$$\frac{\sum m_k Q_k}{\sum m_k} = \frac{\sum m_k (P_k + A)}{\sum m_k} = \frac{\sum m_k P_k}{\sum m_k} + A = C + A.$$

Quer isto dizer que a posição do centro de massa depende unicamente dos pontos P_1, P_2, \dots, P_n e das respectivas massas, mas não da origem. O centro de massa é uma quantidade calculada teoricamente e representa, por assim dizer, um fictício “ponto de equilíbrio” do sistema.

Se as massas estão todas situadas num plano em pontos com coordenadas $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ e se o centro de massa tem coordenadas (\bar{x}, \bar{y}) , a equação vectorial que define C pode ser substituída pelas equações escalares

$$\bar{x} = \frac{\sum m_k x_k}{\sum m_k} \quad \text{e} \quad \bar{y} = \frac{\sum m_k y_k}{\sum m_k}.$$

No numerador da fracção definindo \bar{x} , o termo de ordem k da soma, $m_k x_k$, diz-se o *momento estático* da massa m_k relativamente ao eixo OY . Se uma massa m igual à massa total do sistema, fosse colocada no centro de massa, o seu momento estático relativamente ao eixo OY seria igual ao do sistema

$$m\bar{x} = \sum_{k=1}^n m_k x_k.$$

Quando consideramos sistemas cuja massa total está distribuída em certa região do plano, em vez de estar situada num número finito de pontos discretos, os conceitos de massa, centro de massa, e momento de inércia definem-se por intermédio de integrais, em vez de somas. Por exemplo, consideremos uma lâmina que tenha a forma de uma região plana S . Admitamos a matéria distribuída sobre essa lâmina com uma densidade conhecida. Com isto significamos que existe uma função não negativa f , definida em S e que $f(x, y)$ representa a massa por unidade de área no ponto (x, y) . Se a placa for construída com um material homogéneo, então a densidade é constante. Neste caso a massa total da placa obtem-se efectuando o produto da densidade pela área da placa.

Quando a densidade varia de ponto para ponto utilizamos o integral duplo da densidade como definição da massa total. Por outras palavras, se a função densidade f é integrável sobre S , definimos a massa total $m(S)$ da placa pela equação

$$m(S) = \iint_S f(x, y) \, dx \, dy.$$

O cociente

$$\frac{\text{massa}}{\text{área}} = \frac{\iint_S f(x, y) \, dx \, dy}{\iint_S dx \, dy}$$

chama-se a *densidade média* da placa. Se S é considerada como uma figura geométrica em vez de uma placa, este cociente define o *valor médio* da função f em S . Neste caso não se exige que f seja não negativa.

Por analogia com o caso finito, definimos *centro de massa* da placa como sendo o ponto (\bar{x}, \bar{y}) determinado pelas equações

$$(11.14) \quad \bar{x}m(S) = \iint_S xf(x, y) \, dx \, dy \quad \text{e} \quad \bar{y}m(S) = \iint_S yf(x, y) \, dx \, dy.$$

Os integrais do segundo membro são os momentos da placa relativamente ao eixo OY e OX , respectivamente. Quando a densidade é constante, por exemplo $f(x, y) = c$, o factor c aparecendo em cada um dos membros das equações (11.14) elimina-se e obtemos

$$\bar{x}a(S) = \iint_S x \, dx \, dy \quad \text{e} \quad \bar{y}a(S) = \iint_S y \, dx \, dy,$$

em que $a(S)$ é a área de S . Nesta hipótese o ponto (\bar{x}, \bar{y}) chama-se o *centroide* da placa (ou da região S).

Seja L uma recta no plano da placa, seja $\delta(x, y)$ a distância dum ponto (x, y) de S a L . Então o número I_L definido pela equação

$$I_L = \iint_S \delta^2(x, y) f(x, y) \, dx \, dy$$

chama-se o *momento de inércia* da placa relativamente a L . Quando $f(x, y) = 1$, I_L chama-se o momento de inércia ou *segundo momento* da região S relativamente a L . Os momentos de inércia em relação aos eixos OX e OY representam-se por I_x e I_y , respectivamente, e são dados pelos integrais

$$I_x = \iint_S y^2 f(x, y) \, dx \, dy \quad \text{e} \quad I_y = \iint_S x^2 f(x, y) \, dx \, dy.$$

A soma destes dois integrais chama-se o *momento polar de inércia* I_0 em relação à origem

$$I_0 = I_x + I_y = \iint_S (x^2 + y^2) f(x, y) \, dx \, dy.$$

Nota: A massa e o centro de massa de uma placa são propriedades do corpo e são independentes da localização da origem e das direcções dos eixos coordenados. O momento polar de inércia depende da localização da origem, mas não das direcções escolhidas para os eixos coordenados. Os momentos estáticos e momentos de inércia em relação a OX e OY dependem da localização da origem e da orientação dos eixos. Se uma placa homogénea admite um eixo de simetria, o centroide estará sobre esse eixo. Se existirem dois eixos de simetria o centroide estará sobre o ponto de intersecção. Estas propriedades, facilmente demonstráveis a partir das definições dadas, são muitas vezes úteis no cálculo do centro de massa e momentos de inércia pelas simplificações que introduzem nos cálculos.

EXEMPLO 1. Uma placa delgada de densidade constante c está limitada por duas circunferências concêntricas cujos raios são a e b e o centro é a origem, $0 < b < a$. Calcular o momento polar de inércia.

Resolução. O integral para I_0 é

$$I_0 = c \iint_S (x^2 + y^2) \, dx \, dy,$$

com $S = \{(x, y) \mid b^2 \leq x^2 + y^2 \leq a^2\}$. Para simplificar os cálculos escrevemos o integral como uma diferença de dois integrais

$$I_0 = c \iint_{S(a)} (x^2 + y^2) dx dy - c \iint_{S(b)} (x^2 + y^2) dx dy,$$

onde $S(a)$ e $S(b)$ são discos circulares de raios a e b , respectivamente. Podemos recorrer à integração repetida para calcularmos o integral estendido a $S(a)$ para encontrarmos

$$\iint_{S(a)} (x^2 + y^2) dx dy = 4 \int_0^a \left[\int_0^{\sqrt{a^2 - x^2}} (x^2 + y^2) dy \right] dx = \frac{\pi a^4}{2}.$$

(Omitimos os pormenores dos cálculos devido ao facto do integral poder calcularse mais facilmente recorrendo a coordenadas polares, o que será estudado na secção 11.27.) Portanto

$$I_0 = \frac{\pi c}{2} (a^4 - b^4) = \pi c (a^2 - b^2) \frac{(a^2 + b^2)}{2} = m \frac{a^2 + b^2}{2},$$

com $m = \pi c (a^2 - b^2)$, a massa da placa.

EXEMPLO 2. Determinar o centroide da região plana limitada por um arco de senoide.

Resolução. Tomamos a região S limitada pela curva $y = \sin x$ e com $0 \leq x \leq \pi$. Por simetria, a abcissa \bar{x} do centroide é $\bar{x} = \pi/2$. A ordenada \bar{y} é dada por

$$\bar{y} = \frac{\iint_S y dx dy}{\iint_S dx dy} = \frac{\int_0^\pi [\int_0^{\sin x} y dy] dx}{\int_0^\pi \sin x dx} = \frac{\int_0^\pi \frac{1}{2} \sin^2 x dx}{2} = \frac{\pi}{8}.$$

11.17. Dois teoremas de Pappus

Pappus de Alexandria, que viveu no final do Século II, foi um dos últimos geómetras da Escola de Alexandria de matemáticos gregos. Escreveu um compêndio de oito livros onde recolheu a maior parte dos conhecimentos matemáticos daquele tempo. (Deles se conservam os seis últimos e parte do segundo.) Pappus descobriu um certo número de interessantes propriedades dos centroides, duas das quais se expõem a seguir. A primeira relaciona o centroide de uma região plana com o volume do sólido de revolução obtido por rotação da região em torno de um eixo do seu plano.

Consideremos uma região plana Q , situada entre os gráficos de duas funções contínuas f e g num intervalo (a, b) , com $0 \leq g \leq f$. Seja S o sólido de revolução gerado pela rotação de Q em torno de eixo. Seja $a(Q)$ a área de Q , $v(S)$ o volume de S , e \bar{y} a ordenada do centroide de Q . Quando Q roda para gerar S , o centroide descreve uma circunferência de raio \bar{y} . O teorema de Pappus estabelece que o volume de S é igual ao perímetro desta circunferência multiplicado pela área de Q , isto é,

$$(11.15) \quad v(S) = 2\pi \bar{y} a(Q).$$

Para demonstrar esta fórmula observamos que o volume é dado pelo integral

$$v(S) = \pi \int_a^b [f^2(x) - g^2(x)] dx$$

e que \bar{y} é dado por

$$\bar{y} a(Q) = \iint_Q y dy dx = \int_a^b \left[\int_{g(x)}^{f(x)} y dy \right] dx = \int_a^b \frac{1}{2} [f^2(x) - g^2(x)] dx.$$

Comparando estas duas fórmulas obtemos (11.15).

EXEMPLO 1. Volume dum toro. Seja S o toro gerado pela rotação de um eixo circular Q de raio R em torno de um eixo do seu plano e a uma distância $b > R$ do centro de Q . O volume de S é facilmente calculável pelo teorema de Pappus. Temos $\bar{y} = b$ e $a(Q) = \pi R^2$, pelo que

$$v(S) = 2\pi \bar{y} a(Q) = 2\pi^2 R^2 b.$$

O exemplo seguinte mostra que o teorema de Pappus pode também usarse na determinação de centroides.

EXEMPLO 2. Centroide de um semicírculo. Seja \bar{y} a ordenada do centroide; então

$$Q = \{(x, y) \mid x^2 + y^2 \leq R^2, y \geq 0\}.$$

A área de Q é $\frac{1}{2}\pi R^2$. Quando Q roda em torno do eixo OX gera uma esfera cujo volume é $\frac{4}{3}\pi R^3$. Pelo teorema de Pappus temos

$$\frac{4}{3}\pi R^3 = 2\pi \bar{y} (\frac{1}{2}\pi R^2),$$

onde resulta $\bar{y} = \frac{4R}{3\pi}$.

O outro teorema de Pappus, que apresentamos a seguir, estabelece que o *centroide da reunião de duas regiões planas disjuntas A e B está situado sobre o segmento definido pelos centroides de A e B* . Mais geralmente, sejam A e B duas placas delgadas que são ou disjuntas, ou se intersectam segundo um conjunto de pontos de medida nula. Sejam $m(A)$ e $m(B)$ as massas de A e B e representem C_A e C_B vectores dirigidos da origem para os respectivos centros de massa. Então a união $A \cup B$ tem massa $m(A) + m(B)$ e o seu centro de massa é determinado pelo vector C , onde

$$(11.16) \quad C = \frac{m(A)C_A + m(B)C_B}{m(A) + m(B)}.$$

O cociente que define C é uma combinação linear da forma $aC_A + bC_B$, com a e b escalares não negativos cuja soma vale 1. Uma combinação linear desta forma diz-se uma

combinação convexa de C_A e C_B . A extremidade de C está pois no segmento de recta definido por C_A e C_B .

A fórmula de Pappus (11.16) resulta imediatamente da definição de centro de massa dada em (11.14). A demonstração é deixada ao leitor como exercício. O teorema pode generalizar-se de uma maneira simples à reunião de três ou mais regiões. É particularmente útil na prática quando uma placa de densidade constante é constituída por diferentes partes, cada uma das quais possui simetria geométrica. Determinamos o centroide de cada parte e formamos depois uma adequada combinação convexa para determinarmos o centroide da reunião. No Exercício 21 da Secção seguinte dão-se exemplos.

11.18. Exercícios

Nos exercícios 1 a 8, a região S é limitada por uma ou mais curvas definidas pelas equações dadas. Em cada caso desenhar a região S e determinar as coordenadas \bar{x} e \bar{y} do centroide.

1. $y = x^2$, $x + y = 2$.
2. $y^2 = x + 3$, $y^2 = 5 - x$.
3. $x - 2y + 8 = 0$, $x + 3y + 5 = 0$, $x = -2$, $x = 4$.
4. $y = \sin^2 x$, $y = 0$, $0 \leq x \leq \pi$.
5. $y = \sin x$, $y = \cos x$, $0 \leq x \leq \frac{\pi}{4}$.
6. $y = \log x$, $y = 0$, $1 \leq x \leq a$.
7. $\sqrt{x} + \sqrt{y} = 1$, $x = 0$, $y = 0$.
8. $x^{3/4} + y^{3/4} = 1$, $x = 0$, $y = 0$, no primeiro quadrante.
9. Uma placa delgada é limitada por um arco de parábola $y = 2x - x^2$ e por OX no intervalo $0 \leq x \leq 2$. Determinar a sua massa se a densidade em cada ponto (x, y) é $(1 - y)/(1 + x)$.
10. Determinar o centro de massa de uma placa delgada com a forma de rectângulo $ABCD$, se a densidade em cada ponto é igual ao produto das distâncias do ponto aos dois lados consecutivos AB e AD .

Nos Exercícios 11 a 16, calcular os momentos de inércia I_x e I_y de uma placa delgada S no plano XOY , limitada por uma ou mais curvas definidas pelas equações dadas. Em cada exercício $f(x, y)$ representa a densidade num ponto arbitrário (x, y) de S .

11. $y = \sin^2 x$, $y = -\sin^2 x$, $-\pi \leq x \leq \pi$; $f(x, y) = 1$.
12. $\frac{x}{a} + \frac{y}{b} = 1$, $\frac{x}{c} + \frac{y}{b} = 1$, $y = 0$, $0 < c < a$, $b > 0$; $f(x, y) = 1$.
13. $(x - r)^2 + (y - r)^2 = r^2$, $x = 0$, $y = 0$, $0 \leq x \leq r$, $0 \leq y \leq r$; $f(x, y) = 1$.
14. $xy = 1$, $xy = 2$, $x = 2y$, $y = 2x$, $x > 0$, $y > 0$; $f(x, y) = 1$.
15. $y = e^x$, $y = 0$, $0 \leq x \leq a$; $f(x, y) = xy$.
16. $y = \sqrt{2x}$, $y = 0$, $0 \leq x \leq 2$; $f(x, y) = |x - y|$.
17. Seja S uma placa delgada de massa m e L_0 e L duas rectas paralelas situadas no plano de S , das quais L_0 passa pelo centro de massa de S . Demonstrar o teorema dos eixos paralelos:

$$I_L = I_{L_0} + mh^2,$$

onde h é a distância entre as duas rectas L e L_0 .

(Sugestão: Uma criteriosa escolha dos eixos simplificará a demonstração).

18. O contorno de uma placa é uma elipse de semi-eixos a e b . Seja L a recta do plano da placa passando pelo centro da elipse e fazendo um ângulo α com o eixo de comprimento $2a$. Se a densidade é constante e se a massa da placa é m , provar que o momento de inércia I_L é igual a $\frac{1}{4}m(a^2 \sin^2 \alpha + b^2 \cos^2 \alpha)$.
19. Determinar a distância média de um vértice de um quadrado de lado h aos pontos interiores do quadrado.
20. Seja δ a distância de um ponto arbitrário P , no interior de um círculo de raio r , a um ponto fixo P_0 cuja distância ao centro do círculo é h . Determinar o valor médio da função δ^2 sobre a região definida pelo círculo.
21. Sejam A, B, C os seguintes rectângulos do plano XOY .

$$A = [0, 4] \times [0, 1], \quad B = [2, 3] \times [1, 3], \quad C = [2, 4] \times [3, 4].$$

Usar o teorema de Pappus para determinar o centroide de cada uma das seguintes figuras:

- (a) $A \cup B$. (c) $B \cup C$.
- (b) $A \cup C$. (d) $A \cup B \cup C$.
22. Um triângulo isósceles T tem base 1 e altura h . A base de T coincide com um lado do rectângulo R de base 1 e altura 2. Determinar um valor de h de tal maneira que o centroide de $R \cup T$ esteja sobre o lado comum de R e T .
23. Um triângulo isósceles T tem base $2r$ e altura h . A base de T coincide com o diâmetro do contorno de um semicírculo D de raio r . Determinar a relação que deve existir entre r e h , de tal maneira que o centroide de $T \cup D$ esteja situado no interior do triângulo.

11.19. Teorema de Green no plano

O segundo teorema fundamental do cálculo, para os integrais de linha, estabelece que o integral de linha de um gradiente ∇f ao longo de uma linha unindo dois pontos a e b pode exprimir-se em função dos valores $f(a)$ e $f(b)$. Existe um teorema análogo, em duas dimensões, pelo qual se exprime o integral duplo estendido a uma região plana R como um integral de linha tomado ao longo da curva fechada que constitui a fronteira de R . Este teorema é usualmente referido como o *teorema de Green*. § Pode enunciar-se de várias maneiras; a mais corrente é na forma da identidade:

$$(11.17) \quad \iint_R \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx dy = \oint_C P dx + Q dy.$$

A curva C , que aparece no segundo membro, é o contorno da região R , e o símbolo de integração \oint indica que esta se faz ao longo desse contorno e no sentido directo, como se indica na figura 11.13.

§ Em homenagem a George Green (1793-1841), um matemático inglês que se debruçou sobre as aplicações da matemática à electricidade e magnetismo, fluxo de fluidos e à reflexão e refração da luz e do som. O teorema que ostenta o nome de Green aparece já nos trabalhos de Gauss e Lagrange.

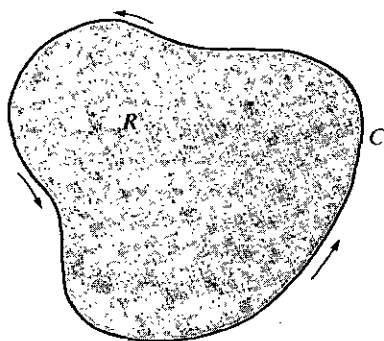


FIG. 11.13. A curva C é o contorno de R , percorrido no sentido directo

São necessários dois tipos de hipóteses para a validade desta identidade. Primeiro, impõem-se condições às funções P e Q para assegurar a existência dos integrais. As mais frequentes são que P e Q sejam continuamente diferenciáveis num conjunto aberto S contendo a região R . Isto implica a continuidade de P e Q em C , bem como a continuidade de $\partial P/\partial y$ e $\partial Q/\partial x$ em R , embora o teorema seja válido sob hipóteses menos restrictivas. Segundo, existem condições de natureza geométrica a serem impostas à região R e à sua fronteira C . A curva C pode ser qualquer *curva fechada, simples, rectificável*. O termo “rectificável” significa, evidentemente, que C tem um comprimento finito. Para explicar o que deve entender-se por curva fechada simples fazemos referência à função vectorial que a define.

Suponhamos que C é descrita por uma função vectorial contínua α definida num intervalo $[a, b]$. Se $\alpha(a) = \alpha(b)$, a curva diz-se *fechada*. Uma curva fechada na qual $\alpha(t_1) \neq \alpha(t_2)$ para todo o par de valores $t_1 \neq t_2$ do intervalo semi-aberto $(a, b]$ diz-se uma curva fechada *simples*. Quer isto dizer que, excepto para os extremos do intervalo $[a, b]$, valores diferentes de t dão lugar a pontos distintos da curva. A circunferência é o exemplo típico da curva fechada simples.

Curvas fechadas simples planas chamam-se habitualmente *curvas de Jordan*, em homenagem a Camille Jordan (1838-1922), um famoso matemático francês que foi dos pioneiros nos conceitos de curvas fechadas simples e de comprimento de arco. Toda a curva de Jordan C decompõe o plano em dois conjuntos abertos conexos disjuntos admitindo a curva C como fronteira comum. Uma dessas regiões é *limitada* e chama-se o *interior* (ou *região interior*) de C . (Um exemplo é a parte sombreada da figura 11.13). A outra é não limitada e chama-se o *exterior* (ou *região exterior*) de C . Para algumas curvas de Jordan familiares tais como circunferências, elipses, ou polígonos elementares, é instructivamente evidente que a curva divide o plano numa região interior e noutra exterior, mas provar que isto é verdadeiro para uma curva de Jordan *qualquer* não é tarefa fácil. Jordan foi o primeiro a referir que esta proposição exige demonstração; o resultado é conhecido pelo *teorema da curva de Jordan*. Até final do século XIX Jordan e outros matemáticos publicaram demonstrações incompletas. Em 1905 o matemático americano Oswald Veblen (1880-1960) estabeleceu a primeira

demonstração completa deste teorema. O teorema de Green é válido sempre que C seja uma curva de Jordan rectificável, e a região R a reunião de C com o seu interior†. Uma vez que não definimos integrais de linha ao longo de curvas arbitrárias rectificáveis, restringimos a nossa discussão aqui a curvas seccionalmente regulares.

Existe outra dificuldade técnica associada com a formulação do teorema de Green. Já observámos que, para a validade da identidade (11.17), a curva C deve ser percorrida no sentido directo. Intuitivamente, isto significa que um indivíduo caminhando ao longo da curva neste sentido, tem sempre a região R à sua esquerda. Mais uma vez, para algumas curvas de Jordan usuais tais como as referidas atrás, o significado da expressão “percorrer a curva no sentido directo” é intuitivamente evidente. Contudo, num tratamento absolutamente rigoroso do teorema de Green ter-se-á que definir esta expressão em termos completamente analíticos, isto é, em termos de função vectorial α que define a curva. Uma definição possível é a esboçada na Secção 11.24.

Tendo assinalado algumas das dificuldades associadas com a formulação do teorema de Green, vamos enunciá-lo na forma geral e indicaremos, de maneira abreviada, porque é verdadeiro para certas regiões particulares. Nesta discussão o significado de “sentido directo” será intuitivo, pelo que o tratamento não será completamente rigoroso.

TEOREMA 11.10. TEOREMA DE GREEN PARA REGIÕES PLANAS LIMITADAS POR CURVAS DE JORDAN SECCIONALMENTE REGULARES. *Sejam P e Q campos escalares continuamente diferenciáveis num conjunto aberto S do plano XOY . Seja C uma curva de Jordan seccionalmente regular, e represente R a união de C com o seu interior. Admita-se que R é um subconjunto de S . Então verifica-se a identidade*

$$(11.18) \quad \iint_R \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx dy = \oint_C P dx + Q dy,$$

onde o integral de linha se torna ao longo de C no sentido directo.

Nota: A identidade em 11.18 é equivalente às duas fórmulas

$$(11.19) \quad \iint_R \frac{\partial Q}{\partial x} dx dy = \oint_C Q dy$$

e

$$(11.20) \quad - \iint_R \frac{\partial P}{\partial y} dx dy = \oint_C P dx.$$

Com efeito, se ambas são verdadeiras, somando-as membro a membro deduz-se (11.18). Inversamente, se (11.18) é verdadeiro podemos obter (11.19) e (11.20) como casos particulares fazendo $P = 0$ e $Q = 0$, respectivamente.

† Na obra do autor *Mathematical Analysis*, (Trad. espanhola *Análisis Matemático*) no capítulo 10 encontra-se uma demonstração do teorema de Green para regiões com esta generalidade.

Demonstração para regiões particulares. Vamos provar (11.20) para uma região R do Tipo I. Uma tal região é da forma

$$R = \{(x, y) \mid a \leq x \leq b \quad \text{e} \quad f(x) \leq y \leq g(x)\},$$

onde f e g são contínuas em $[a, b]$ com $f \leq g$. A fronteira C de R é formada por quatro partes, um arco inferior C_1 (o gráfico de f), outro superior C_2 (o gráfico de g), e dois

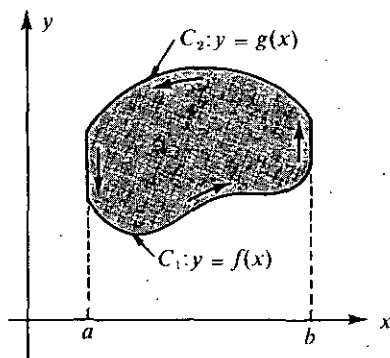


FIG. 11.14 Demonstração do teorema de Green para uma região particular.

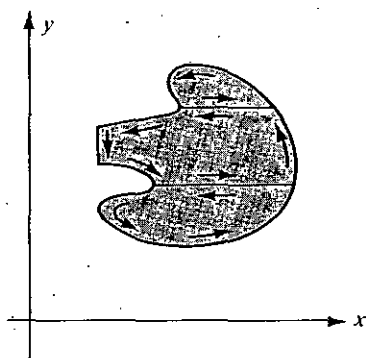


FIG. 11.15 Demonstração do teorema de Green para uma região mais geral.

segmentos de recta paralelos a OY , percorridos no sentido indicado pelas setas (figura 11.14).

Calculemos em primeiro lugar o integral duplo $-\iint_R (\partial P / \partial y) \, dx \, dy$ por integração repetida. Integrando em primeiro lugar com respeito a y temos

$$\begin{aligned} (11.21) \quad -\iint_R \frac{\partial P}{\partial y} \, dx \, dy &= -\int_a^b \left[\int_{f(x)}^{g(x)} \frac{\partial P}{\partial y} \, dy \right] dx = \int_a^b \left[\int_{g(x)}^{f(x)} \frac{\partial P}{\partial y} \, dy \right] dx \\ &= \int_a^b P[x, f(x)] \, dx - \int_a^b P[x, g(x)] \, dx. \end{aligned}$$

Por outro lado, o integral curvilíneo $\int_C P \, dx$ pode escrever-se:

$$\int_C P \, dx = \int_{C_1} P \, dx + \int_{C_2} P \, dx,$$

uma vez que o integral de linha, ao longo de cada segmento paralelo a OY , é zero. Para calcular o integral ao longo de C_1 usamos a representação vectorial $\alpha(t) = ti + f(t)j$ para obtermos

$$\int_{C_1} P dx = \int_a^b P[t, f(t)] dt.$$

Para calcular o integral ao longo de C_2 usamos a representação $\alpha(t) = ti + g(t)j$ e tomamos em conta a inversão do sentido para obtermos

$$\int_{C_2} P dx = -\int_a^b P[t, g(t)] dt.$$

Portanto temos

$$\int_C P dx = \int_a^b P[t, f(t)] dt - \int_a^b P[t, g(t)] dt.$$

Comparando esta equação com a fórmula (11.21) obtemos (11.20).

Um raciocínio semelhante permitirá demonstrar (11.19) para regiões do Tipo II. Obtem-se assim uma demonstração do teorema de Green quer para regiões do Tipo I, quer do Tipo II. Uma vez feito isto, o teorema pode ser demonstrado para aquelas regiões R que podem decompor-se num número finito de regiões daqueles dois tipos. Com essa finalidade efectuem-se cortes transversais, como se representa na figura 11.15, aplicam-se o teorema a cada subregião e adicionam-se os resultados. Os integrais de linha ao longo das secções transversais anulam-se aos pares, como se sugere na figura, e a soma dos integrais de linha ao longo das fronteiras das subregiões é igual ao integral de linha ao longo da fronteira de R .

11.20. Algumas aplicações do teorema de Green

Os exemplos que apresentamos a seguir ilustram algumas aplicações do teorema de Green.

EXEMPLO 1. Calcular, por intermédio do teorema de Green, o trabalho efectuado pelo campo de forças $f(x, y) = (y + 3x)i + (2y - x)j$ actuando sobre uma partícula obrigando-a a descrever uma vez a elipse $4x^2 + y^2 = 4$ no sentido directo.

Resolução. O trabalho é igual a $\int_C P dx + Q dy$, sendo $P = y + 3x$, $Q = 2y - x$ e C a elipse. Uma vez que $\partial Q / \partial x - \partial P / \partial y = -2$, o teorema de Green dá-nos

$$\int_C P dx + Q dy = \iint_R (-2) dx dy = -2a(R),$$

em que $a(R)$ é a área da região interior à elipse. Porque a elipse tem semi-eixos $a = 1$ e $b = 2$, a sua área é $\pi ab = 2\pi$ e o valor do integral de linha é -4π .

EXEMPLO 2. Calcular o integral de linha $\int_C (5 - xy - y^2)dx - (2xy - x^2)dy$, sendo C o quadrado de vértices $(0, 0)$, $(1, 0)$, $(1, 1)$, $(0, 1)$, percorrido no sentido directo.

Resolução. Neste caso $P = 5 - xy - y^2$, $Q = x^2 - 2xy$, e $\partial Q/\partial x - \partial P/\partial y = 3x$. Logo, pelo teorema de Green, temos

$$\int_C P dx + Q dy = 3 \iint_R x dx dy = 3\bar{x},$$

representando \bar{x} a abscissa do centroide do quadrado. Visto que \bar{x} é obviamente $\frac{1}{2}$, o valor do integral de linha é $\frac{3}{2}$.

EXEMPLO 3. A área expressa como um integral de linha. O integral duplo para a área $a(R)$ da região R pode escrever-se

$$a(R) = \iint_R dx dy = \iint_R \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx dy,$$

em que P e Q são tais que $\partial Q/\partial x - \partial P/\partial y = 1$. Por exemplo, podemos tomar $Q(x, y) = \frac{1}{2}x$ e $P(x, y) = -\frac{1}{2}y$. Se R é uma região cuja fronteira é uma curva de Jordan C , podemos aplicar o teorema de Green para exprimir $a(R)$ como um integral de linha,

$$a(R) = \int_C P dx + Q dy = \frac{1}{2} \int_C -y dx + x dy.$$

Se a curva fronteira C se define parametricamente por

$$x = X(t), \quad y = Y(t), \quad a \leq t \leq b,$$

o integral de linha para a área escreve-se

$$a(R) = \frac{1}{2} \int_a^b \{-Y(t)X'(t) + X(t)Y'(t)\} dt = \frac{1}{2} \int_a^b \begin{vmatrix} X(t) & Y(t) \\ X'(t) & Y'(t) \end{vmatrix} dt.$$

11.21 Uma condição necessária e suficiente para que um campo vectorial bidimensional seja um gradiente

Seja $f(x, y) = P(x, y)i + Q(x, y)j$ um campo vectorial que é continuamente diferenciável num conjunto aberto S do plano. Se f é um gradiente em S temos

$$(11.22) \quad \frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}$$

em todo S . Por outras palavras, a condição (11.22) é necessária para que f seja um gradiente. Como já foi observado, esta condição não é suficiente. Por exemplo, o campo vectorial

$$f(x, y) = \frac{-y}{x^2 + y^2} \mathbf{i} + \frac{x}{x^2 + y^2} \mathbf{j}$$

verifica (11.22) em todo o conjunto $S = \mathbb{R}^2 - \{(0, 0)\}$, mas f não é um gradiente em S . No teorema 10.9 provou-se que a condição (11.22) é uma condição necessária e suficiente para que f seja um gradiente em S se S é convexo. Com auxílio do teorema de Green podemos generalizar este resultado a uma classe mais geral de conjuntos planos, a dos conjuntos *simplesmente conexos*. Estes definem-se como a seguir se indica.

DEFINIÇÃO DE UM CONJUNTO PLANO SIMPLESMENTE CONEXO. *Seja S um conjunto conexo aberto no plano. Então S diz-se simplesmente conexo se, para cada curva de Jordan C situada em S , o interior de C é também um subconjunto de S .*

Uma coroa circular (conjunto de pontos situados entre duas circunferências concêntricas) não é simplesmente conexa porque toda a região interior a uma circunferência concêntrica com as anteriores e cujo raio esteja compreendido entre os daquelas não é um subconjunto da coroa. Intuitivamente falando, dizemos que S é simplesmente conexo quando não tem “buracos”. Outra maneira de descrever a conexão simples consiste em dizer que nela uma curva C_1 de S ligando dois quaisquer pontos pode ser continuamente deformada até outra curva qualquer C_2 de S unindo esses dois pontos, com todas as curvas intermédias durante a deformação permanecendo completamente em S . Uma outra definição, a qual pode mostrar-se ser equivalente à anterior, estabelece que um conjunto conexo aberto S é simplesmente conexo se o seu complemento (relativo a todo o plano) não é conexo. Por exemplo, uma coroa circular não é simplesmente conexa porque o seu complemento é não conexo. Um conjunto aberto conexo que não seja simplesmente conexo chama-se *multiplamente conexo*.

TEOREMA 11.11. *Seja $f(x, y) = P(x, y)\mathbf{i} + Q(x, y)\mathbf{j}$ um campo vectorial continuamente diferenciável num conjunto S simplesmente conexo aberto no plano. Então f é um gradiente em S se e só se*

$$(11.23) \quad \frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x} \quad \text{em todos os pontos de } S.$$

Demonstração. Como já se fez notar, a condição (11.23) é necessária para que f seja um gradiente. Vamos agora provar que também é uma condição suficiente.

Pode demonstrar-se que em qualquer conjunto plano conexo aberto S , todo par de pontos a e x pode unir-se mediante uma linha poligonal em escada simples, isto é, por uma escala poligonal cujos lados são paralelos aos eixos coordenados e que não se intersectam. Se o integral de linha de f , de a a x , tem o mesmo valor para cada linha

polihonal em escada simples de S unindo a a x , então o mesmo raciocínio empregado para demonstrar o teorema 10.4 prova que f é um gradiente em S . Por conseguinte, necessitamos verificar somente que o integral de linha de f , desde a até x , tem o mesmo valor para toda a linha poligonal em escada em S unindo a a x .

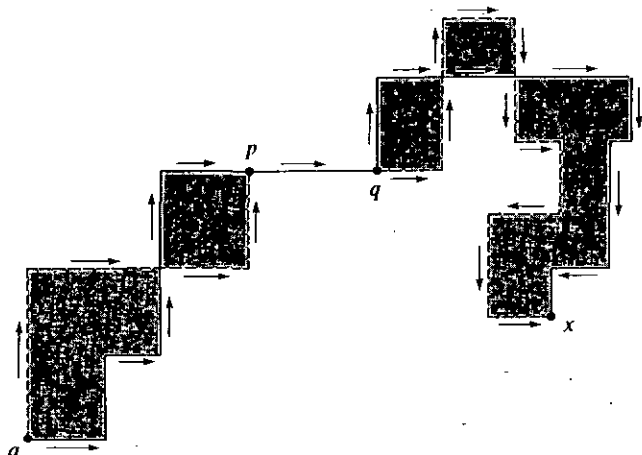


FIG. 11.16. Independência da linha numa região simplesmente conexa.

Sejam C_1 e C_2 duas linhas poligonais em escada simples de S unindo a a x . Algumas partes destas linhas poligonais podem coincidir ao longo de certos segmentos de recta. As restantes partes intersectar-se-ão quando muito um número finito de vezes, e formarão fronteiras de um número finito de regiões poligonais, seja R_1, \dots, R_m . Uma vez que S se supõe ser simplesmente conexa, cada uma das regiões R_k é um subconjunto de S . Na figura 11.16 apresenta-se um exemplo. A linha a cheio representa C_1 e a linha a tracejado representa C_2 e as partes sombreadas representam R_1, \dots, R_m . (Ao longo do segmento pq essas duas linhas poligonais coincidem).

Observemos agora que o integral de linha de f desde a a x ao longo de C_1 , mais o integral de x até a ao longo de C_2 é zero porque o integral ao longo da linha fechada é uma soma de integrais tomados sobre os segmentos comuns a C_1 e C_2 mais os integrais tomados ao longo das fronteiras das regiões R_k . Os integrais calculados ao longo dos segmentos comuns anulam-se dois a dois, uma vez que cada segmento comum é percorrido duas vezes, em sentidos opostos, e a sua soma é zero. O integral sobre a fronteira Γ_k de cada região R_k é também zero porque, pelo teorema de Green, podemos escrever

$$\int_{\Gamma_k} P dx + Q dy = \pm \iint_{R_k} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx dy,$$

e o integrando do integral duplo é zero devido à hipótese $\partial P/\partial y = \partial Q/\partial x$. Resulta que o integral de a a x , ao longo de C_1 , é igual ao calculado ao longo de C_2 . Como já se observou, isto implica que f é um gradiente em S .

11.22. Exercícios

1. Aplicar o teorema de Green para calcular o integral de linha $\oint_C y^2 dx + x dy$ quando
 - (a) C é o quadrado com vértices $(0, 0)$, $(2, 0)$, $(2, 2)$, $(0, 2)$.
 - (b) C é o quadrado com vértices $(\pm 1, \pm 1)$.
 - (c) C é o quadrado com vértices $(\pm 2, 0)$, $(0, \pm 2)$.
 - (d) C é o círculo de raio 2 e centro na origem.
 - (e) C tem a equação vectorial $\alpha(t) = 2 \cos t \mathbf{i} + 2 \sin t \mathbf{j}$, $0 \leq t \leq 2\pi$.
2. Se $P(x, y) = xe^{-y^2}$ e $Q(x, y) = -x^2y e^{-y^2} + 1/(x^2 + y^2)$, calcular o integral de linha $\oint_C P dx + Q dy$ ao longo da fronteira do quadrado de lado $2a$ determinado pelas desigualdades $|x| \leq a$ e $|y| \leq a$.
3. Seja C uma curva fechada simples no plano XOY e I_z o momento de inércia (em relação ao eixo OZ) da região interior a C . Mostrar que existe um inteiro n tal que

$$nI_z = \oint_C x^3 dy - y^3 dx.$$

4. Dados dois campos escalares u e v , continuamente diferenciáveis num conjunto aberto que contém o disco circular R cuja fronteira é a circunferência $x^2 + y^2 = 1$, definem-se dois campos vectoriais f e g do modo seguinte:

$$f(x, y) = v(x, y)\mathbf{i} + u(x, y)\mathbf{j}, \quad g(x, y) = \left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right)\mathbf{i} + \left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right)\mathbf{j}.$$

Determinar o valor do integral duplo $\iint_R f \cdot g \, dx \, dy$ se é sabido que sobre a fronteira de R se tem $u(x, y) = 1$ e $v(x, y) = y$.

5. Se f e g são continuamente diferenciáveis num conjunto conexo aberto S do plano, provar que $\oint_C f \nabla g \cdot d\alpha = -\oint_C g \nabla f \cdot d\alpha$ para toda a curva de Jordán C seccionalmente regular em S .
6. Sejam u e v campos escalares admitindo derivadas parciais de primeira e segunda ordens continuas num conjunto aberto conexo S do plano. Seja R uma região de S cuja fronteira é uma curva de Jordan C seccionalmente regular. Provar que:

$$(a) \oint_C uv \, dx + uv \, dy = \iint_R \left\{ v \left(\frac{\partial u}{\partial x} - \frac{\partial u}{\partial y} \right) + u \left(\frac{\partial v}{\partial x} - \frac{\partial v}{\partial y} \right) \right\} dx \, dy.$$

$$(b) \frac{1}{2} \oint_C \left(v \frac{\partial u}{\partial x} - u \frac{\partial v}{\partial x} \right) dx + \left(u \frac{\partial v}{\partial y} - v \frac{\partial u}{\partial y} \right) dy = \iint_R \left(u \frac{\partial^2 v}{\partial x \partial y} - v \frac{\partial^2 u}{\partial x \partial y} \right) dx \, dy.$$

Derivadas normais. Na secção 10.7 definimos integrais de linha a respeito do comprimento de arco de tal modo que é válida a seguinte igualdade:

$$\int_C P \, dx + Q \, dy = \int_C f \cdot T \, ds,$$

sendo $f = P\mathbf{i} + Q\mathbf{j}$ e T o vector unitário tangente a C . (O produto escalar $f \cdot T$ dá a componente tangencial de f ao longo de C .) Se C é uma curva de Jordan definida por uma função α continuamente diferenciável, seja $\alpha(t) = X(t)\mathbf{i} + Y(t)\mathbf{j}$, o vector unitário normal exterior \mathbf{n} a C define-se pela equação

$$\mathbf{n}(t) = \frac{1}{\|\alpha'(t)\|} (Y'(t)\mathbf{i} - X'(t)\mathbf{j})$$

sempre que $\|\alpha'(t)\| \neq 0$. Se φ é um campo escalar com gradiente $\nabla\varphi$ em C , a derivada normal $\partial\varphi/\partial n$ define-se em C mediante a equação

$$\frac{\partial\varphi}{\partial n} = \nabla\varphi \cdot \mathbf{n}.$$

Esta é, evidentemente, a derivada direccional de φ na direcção de \mathbf{n} . Estes conceitos aparecem nos restantes exercícios desta secção.

7. Se $f = Qi - Pj$, mostrar que

$$\int_C P dx + Q dy = \int_C f \cdot \mathbf{n} ds.$$

(O produto escalar $f \cdot \mathbf{n}$ dá a componente normal de f ao longo de C .)

8. Sejam f e g campos escalares admitindo derivadas parciais contínuas de primeira e segunda ordem num conjunto aberto S no plano. Designese por R uma região (em S) cuja fronteira é uma curva C de Jordan seccionalmente regular. Provar as seguintes identidades, com $\nabla^2 u = \partial^2 u / \partial x^2 + \partial^2 u / \partial y^2$.

$$(a) \oint_C \frac{\partial g}{\partial n} ds = \iint_R \nabla^2 g dx dy.$$

$$(b) \oint_C f \frac{\partial g}{\partial n} ds = \iint_R (f \nabla^2 g + \nabla f \cdot \nabla g) dx dy.$$

$$(c) \oint_C \left(f \frac{\partial g}{\partial n} - g \frac{\partial f}{\partial n} \right) ds = \iint_R (f \nabla^2 g - g \nabla^2 f) dx dy.$$

A identidade (c) é conhecida por fórmula de Green; ela mostra que

$$\oint_C f \frac{\partial g}{\partial n} ds = \oint_C g \frac{\partial f}{\partial n} ds$$

com f e g harmónicas em R (isto é, quando $\nabla^2 f = \nabla^2 g = 0$ em R)

9. Supor que a equação diferencial

$$P(x, y) dx + Q(x, y) dy = 0$$

admite um factor integrante $u(x, y)$ que nos permite obter uma família de soluções a um parâmetro da forma $\varphi(x, y) = C$. Se o declive da curva $\varphi(x, y) = C$ em (x, y) é $\tan \theta$, o vector unitário \mathbf{n} é

$$\mathbf{n} = \sin \theta \mathbf{i} - \cos \theta \mathbf{j}.$$

Existe um campo escalar $g(x, y)$ tal que a derivada normal de ϕ é dada pela fórmula

$$\frac{\partial \phi}{\partial n} = \mu(x, y)g(x, y),$$

onde $\partial \phi / \partial n = \nabla \phi \cdot \mathbf{n}$. Determinar uma fórmula explícita para $g(x, y)$ em função de $P(x, y)$ e $Q(x, y)$.

★ 11.23 Teorema de Green para regiões multiplamente conexas

O teorema de Green pode generalizar-se para se aplicar a certas regiões multiplamente conexas.

TEOREMA 11.12. TEOREMA DE GREEN PARA REGIÕES MULTIPLAMENTE CONEXAS. Sejam C_1, \dots, C_n , n curvas de Jordan seccionalmente regulares gozando das seguintes propriedades:

- (a) Duas quaisquer dessas curvas não se intersectam.
- (b) As curvas C_2, \dots, C_n estão todas situadas no interior de C_1 .
- (c) A curva C_i está no interior da curva C_j para $i \neq j, i > 1, j > 1$.

Seja R a região formada pela união de C_1 com a porção do interior de C_1 que não é interior a qualquer das curvas C_2, C_3, \dots, C_n . (Na figura 11.17 está representado um exemplo de tal região). Sejam P e Q continuamente diferenciáveis num conjunto aberto S contendo R . Então é válida a seguinte identidade:

$$(11.24) \quad \iint_R \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx dy = \oint_{C_1} (P dx + Q dy) - \sum_{k=2}^n \oint_{C_k} (P dx + Q dy).$$

O teorema pode demonstrar-se recorrendo à introdução de secções transversais que transformam R na união de um número finito de regiões simplesmente conexas limitadas por curvas de Jordan. Aplica-se o teorema de Green a cada parte separadamente.

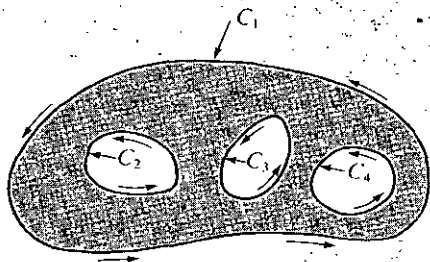


FIG. 11.17. Uma região multiplamente conexa.

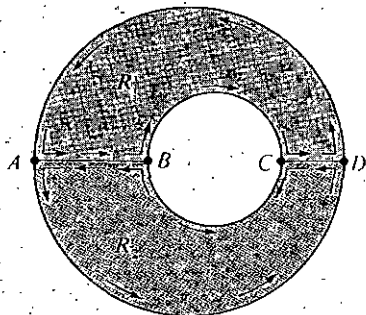


FIG. 11.18. Demonstração do teorema de Green para uma região multiplamente conexa.

te, adicionando-se depois os resultados. Vamos exemplificar como pode efectuar-se a demonstração quando $n = 2$. Por indução pôde demonstrar-se o caso mais geral com um número n de curvas.

A ideia da demonstração quando $n = 2$ está figurada com o exemplo da figura 11.18; C_1 e C_2 são duas circunferências, sendo C_1 a maior. Consideremos as secções AB e CD , como se indica na figura. Seja K_1 a curva de Jordan formada pela metade superior de C_2 , a metade superior de C_1 e os segmentos AB e CD . Seja K_2 a curva de Jordan formada pela metade inferior de C_1 , a metade inferior de C_2 e os dois segmentos AB e CD . Apliquemos agora o teorema de Green a cada uma das regiões limitadas por K_1 e K_2 e adicionemos as duas identidades obtidas. Os integrais de linha ao longo das secções transversais anulam-se (visto cada tal secção ser percorrida duas vezes em sentidos contrários), resultando então a igualdade

$$\iint_R \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx dy = \oint_{C_1} (P dx + Q dy) - \oint_{C_2} (P dx + Q dy).$$

O sinal menos aparece devido ao sentido segundo o qual é percorrida a curva C_2 . Ora a igualdade anterior corresponde a (11.24) quando $n = 2$.

Para uma região simplesmente conexa, a condição $\partial P/\partial y = \partial Q/\partial x$ implica que o integral de linha $\int P dx + Q dy$ é independente do caminho ao longo do qual se considere (teorema 11.11). Como já fizemos notar, se S é não simplesmente conexa, a condição $\partial P/\partial y = \partial Q/\partial x$ não implica necessariamente independência do integral em relação à linha ao longo do qual se calcula. Contudo, neste caso existe um substituto para a condição de independência referida que pode deduzir-se do teorema 11.12.

TEOREMA 11.13. INVARIÂNCIA DE UM INTEGRAL DE LINHA FACE À DEFORMAÇÃO DA LINHA DE INTEGRAÇÃO. *Sejam P e Q continuamente diferenciáveis num conjunto conexo aberto S do plano e admita-se que $\partial P/\partial y = \partial Q/\partial x$ em todo S . Sejam C_1 e C_2 duas curvas de Jordan seccionalmente regulares situadas em S e satisfazendo às seguintes condições:*

- C_2 está no interior de C_1 .
- Os pontos interiores a C_1 que são exteriores a C_2 pertencem a S . (A figura 11.19 representa um exemplo.)

Então tem-se

$$(11.25) \quad \oint_{C_1} P dx + Q dy = \oint_{C_2} P dx + Q dy$$

sendo ambas as curvas percorridas no mesmo sentido.

Demonstração. Sob as condições estabelecidas, a equação (11.24) é aplicável quando $n = 2$. A região R consiste daqueles pontos situados entre as curvas C_1 e C_2 e das próprias curvas. Visto que $\partial P/\partial y = \partial Q/\partial x$ em S , o primeiro membro da equação (11.24) é zero e obtemos (11.25).

O teorema 11.13 é por vezes referido como afirmando que se $\partial P/\partial y = \partial Q/\partial x$ em S , o valor do integral de linha ao longo de uma curva fechada simples, seccionalmente regular, pertencente a S , permanece inalterado se a curva se deforma até coincidir com outra curva simples, fechada, seccionalmente regular de S , desde que todas

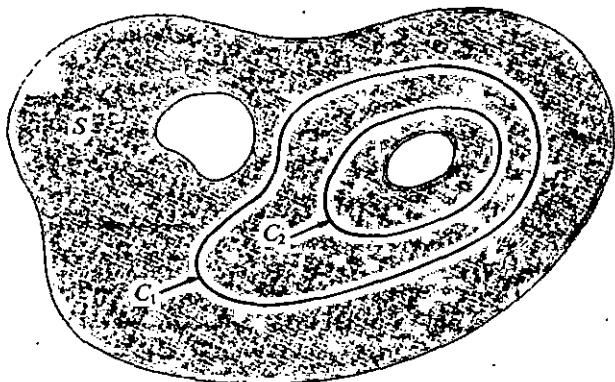


FIG. 11.19. Invariância do integral de linha face à deformação da linha de integração.

as curvas intermediárias que se vão obtendo por deformação da inicial permaneçam em S . O conjunto S supõe-se ser aberto e conexo—não necessariamente simplesmente conexo.

★ 11.24. O número de giros

Vimos que o valor de um integral de linha depende muitas vezes da curva ao longo da qual se efectua a integração e também do sentido segundo o qual a curva é descrita. Por exemplo, a identidade no teorema de Green exige que o integral de linha seja calculado no sentido directo. Num estudo absolutamente rigoroso do teorema de Green será necessário descrever analiticamente o significado da expressão “percorrer uma curva fechada no sentido directo”. Para algumas curvas particulares isto pode conseguir-se por convenções específicas a respeito da função vectorial α que define a curva. Por exemplo, a função vectorial α definida no intervalo $[0, 2\pi]$ pela equação

$$(11.26) \quad \alpha(t) = (a \cos t + x_0)i + (a \sin t + y_0)j$$

define uma circunferência de raio a com centro em (x_0, y_0) . Esta função particular diz-se descrever a circunferência no sentido *positivo* (ou *directo*) (*contrário ao do movimento dos ponteiros do relógio*). Por outro lado, se substituirmos t por $-t$ no segundo membro de (11.26) obtemos uma nova função da qual se diz descrever a circunferência no sentido *negativo* ou *inverso* (*o sentido do movimento dos ponteiros do relógio*). Demos assim uma descrição analítica completa de *sentido positivo* e *sentido negativo*

para a circunferência. Todavia, já não é tão simples definir a mesma ideia para uma curva fechada *arbitrária*. Para curvas seccionalmente regulares tal será possível fazer-se pela introdução do conceito do *número de giros*, um artifício analítico que nos dá um método matematicamente preciso de contar o número de vezes que um raio vector α "gira em torno" de um dado ponto, quando vai descrevendo uma dada curva fechada. Nesta Secção vamos descrever, resumidamente, um método de introdução do número de giros, para em seguida indicarmos como pode ser usado para atribuir sentidos positivos e negativos a curvas fechadas.

Seja C uma curva plana fechada, seccionalmente regular, descrita por uma função vectorial α definida num intervalo $[a, b]$, por exemplo

$$\alpha(t) = X(t)i + Y(t)j \quad \text{se } a \leq t \leq b.$$

Seja $P_0 = (x_0, y_0)$ um ponto não pertencente a C . Então o número de giros de α , em relação ao ponto P_0 , representa-se por $W(\alpha; P_0)$; define-se como sendo o valor do seguinte integral:

$$(11.27) \quad W(\alpha; P_0) = \frac{1}{2\pi} \int_a^b \frac{[X(t) - x_0]Y'(t) - [Y(t) - y_0]X'(t)}{[X(t) - x_0]^2 + [Y(t) - y_0]^2} dt.$$

Este é o mesmo que o integral de linha

$$(11.28) \quad \frac{1}{2\pi} \oint_C \frac{-(y - y_0) dx + (x - x_0) dy}{(x - x_0)^2 + (y - y_0)^2}.$$

Pode demonstrar-se que o valor deste integral é sempre um *inteiro*, positivo, negativo ou nulo. Além disso, se C é uma curva de *Jordan* (curva fechada *simples*) este inteiro é zero se P_0 é *exterior* a C e toma os valores $+1$ ou -1 se P_0 é *interior* a C . (Ver figura 11.20.) Consequentemente, $W(\alpha; P_0)$ é $+1$ ou -1 para *todo* o ponto P_0 interior a

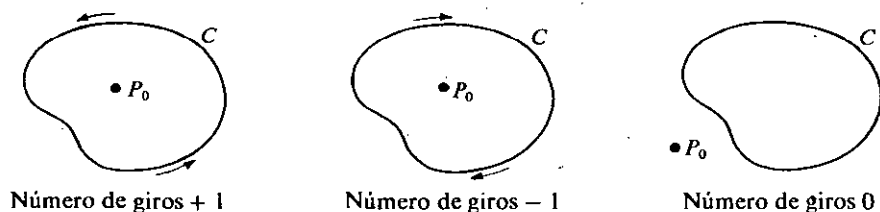


FIG. 11.20. Valores possíveis do número de giros de uma curva de Jordan C com respeito ao ponto P_0 .

C . Tal facto permite-nos definir orientações positivas e negativas para C do modo seguinte: Se o número de giros, $W(\alpha, P_0)$, é $+1$ para todo ponto P_0 interior a C dizemos que α define C no sentido *positivo* ou *directo*. Se o número de giro é -1 dizemos

que α define C no sentido *negativo* ou *inverso*. [um exemplo do integral (11.28) com $x_0 = y_0 = 0$ foi encontrado no Exemplo 2 da Secção 10.16].

Para demonstrarmos que o integral que define o número de giros é sempre $+1$ ou -1 , para uma curva simples fechada contendo no seu interior o ponto (x_0, y_0) , recorreremos ao teorema 11.13. Seja S uma região conexa, aberta, formada por todos os pontos do plano excepto (x_0, y_0) . O integral de linha (11.28) pode então escrever-se na forma $\int_C P dx + Q dy$, e é fácil verificar que $\partial P / \partial y = \partial Q / \partial x$ em toda a região S . Por conseguinte, se (x_0, y_0) é interior a C , o teorema 11.13 diz-nos que podemos substituir a curva C por uma circunferência com centro em (x_0, y_0) sem mudar o valor do integral. Verificamos seguidamente que para uma circunferência o integral que define o número de giros é $+1$ ou -1 , dependendo o valor do facto da circunferência ser descrita positivamente ou negativamente. Para uma circunferência orientada positivamente podemos usar a representação dada em (11.26). Nesse caso temos

$$X(t) = a \cos t + x_0, \quad Y(t) = a \sin t + y_0,$$

e o integrando em (11.27) é identicamente igual a 1, onde resulta

$$W(\alpha; P_0) = \frac{1}{2\pi} \int_0^{2\pi} 1 dt = 1.$$

Por um raciocínio análogo encontramos que o integral vale -1 quando C é descrita negativamente. Está assim demonstrado que o número de giros é $+1$ ou -1 para uma curva fechada simples que circunde o ponto (x_0, y_0) .

★ 11.25. Exercícios

1. Seja $S = \{(x, y) | x^2 + y^2 > 0\}$ e seja ainda

$$P(x, y) = \frac{y}{x^2 + y^2}, \quad Q(x, y) = \frac{-x}{x^2 + y^2}$$

se $(x, y) \in S$. Seja C uma curva de Jordan seccionalmente regular em S .

(a) Se $(0, 0)$ é interior a C , provar que o integral de linha $\int_C P dx + Q dy$ tem o valor $\pm 2\pi$, e explicar quando é que ocorre o sinal $+$.

(b) Calcular o integral de linha $\int_C P dx + Q dy$ quando $(0, 0)$ é exterior a C .

2. Se $r = xi + yj$ e $r = \|r\|$, seja

$$f(x, y) = \frac{\partial(\log r)}{\partial y} i - \frac{\partial(\log r)}{\partial x} j$$

para $r > 0$. Represente C uma curva de Jordan seccionalmente regular situado na coroa circular $1 < x^2 + y^2 < 25$; determinar todos os valores possíveis do integral de linha de f ao longo de C .

3. Uma região plana conexa com um só "buraco" diz-se *duplamente conexa*. (A coroa $1 < x^2 + y^2 < 25$ constitui um exemplo). Se P e Q são continuamente diferenciáveis numa região duplamente conexa R , e se $\partial P / \partial y = \partial Q / \partial x$ em toda a região R , quantos valores distintos são possíveis para os integrais de linha $\int_C P dx + Q dy$, tomados sobre curvas de Jordan seccionalmente regulares e situadas em R ?
4. Resolva o Exercício 3 para regiões triplamente conexas, isto é, para regiões planas conexas só com dois buracos.
5. Sejam P e Q dois campos escalares com derivadas contínuas que satisfazem a $\partial P / \partial y = \partial Q / \partial x$ em todo o plano, excepto em três pontos. Sejam C_1 , C_2 e C_3 três circunferências de centros nesses três pontos, tal como se indica na fig. 11.21; e seja $I_k = \oint_{C_k} P dx + Q dy$. Suponhamos que $I_1 = 12$, $I_2 = 10$, $I_3 = 15$.
- (a) Determinar o valor de $\int_C P dx + Q dy$, onde C é a figura em forma de oito representada na figura.
- (b) Traçar outra curva fechada Γ ao longo da qual $\int \Gamma P dx + Q dy = 1$. Indicar no desenho qual o sentido em Γ é percorrida.
- (c) Se $I_1 = 12$, $I_2 = 9$ e $I_3 = 15$, mostrar que não existe nenhuma curva fechada Γ ao longo da qual $\int \Gamma P dx + Q dy = 1$.

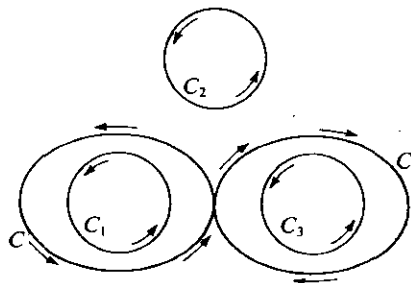


FIG. 11.21. Exercício 5.

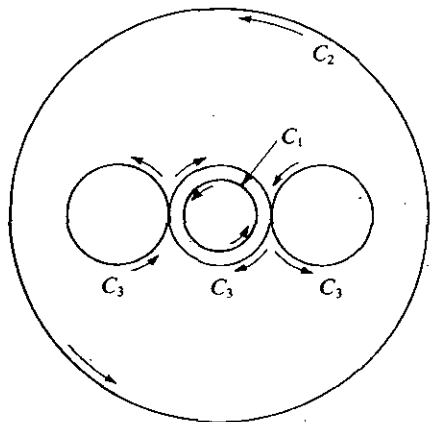


FIG. 11.22. Exercício 6.

6. Seja $I_k = \oint_{C_k} P dx + Q dy$, onde

$$P(x, y) = -y \left[\frac{1}{(x-1)^2 + y^2} + \frac{1}{x^2 + y^2} + \frac{1}{(x+1)^2 + y^2} \right]$$

e

$$Q(x, y) = \frac{x-1}{(x-1)^2 + y^2} + \frac{x}{x^2 + y^2} + \frac{x+1}{(x+1)^2 + y^2}.$$

Na figura 11.22, C_1 é a circunferência menor, $x^2 + y^2 = \frac{1}{4}$ (com sentido positivo), C_2 é a maior, $x^2 + y^2 = 4$ (com sentido negativo), e C_3 é a curva formada por três circunferências intermédias $(x-1)^2 + y^2 = \frac{1}{4}$, $x^2 + y^2 = \frac{1}{4}$, e $(x+1)^2 + y^2 = \frac{1}{4}$ traçadas como se indicam. Se $I_2 = 6\pi$ e $I_3 = 2\pi$, determina o valor de I_1 .

11.26. Mudança de variáveis num integral duplo

Na teoria da integração a uma dimensão, o método de substituição permite-nos frequentemente calcular integrais complicados transformando-os noutros mais simples, ou noutros tipos que podem ser mais facilmente calculados. O método baseia-se na fórmula

$$(11.29) \quad \int_a^b f(x) dx = \int_c^d f[g(t)]g'(t) dt,$$

onde $a = g(c)$ e $b = g(d)$. Demonstrámos esta fórmula (no Volume I) sob a hipótese de que g admite derivada continua num intervalo $[c, d]$ e que f é continua no conjunto de valores que toma $g(t)$ quando t varia no intervalo $[c, d]$.

Existe uma fórmula análoga a (11.29) para o caso bidimensional, chamada a fórmula de mudança de variáveis num integral duplo. Por seu intermédio transforma-se um integral da forma $\int_S f(x, y) dx dy$, estendido a uma região S do plano XOY , noutro integral duplo $\int_T F(u, v) du dv$, estendido a uma nova região T do plano UOV . É a relação entre as regiões S e T e os integrandos $f(x, y)$ e $F(u, v)$ que vamos passar a analisar. O método de substituição para os integrais duplos é mais complicado que

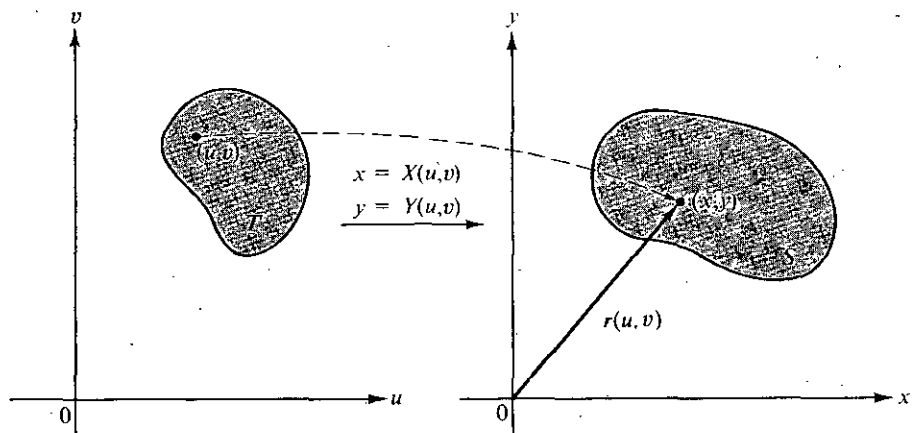


FIG. 11.23. Aplicação definida pela equação vectorial $r(u, v) = X(u, v)\mathbf{i} + Y(u, v)\mathbf{j}$.

no caso unidimensional, porque existem duas substituições a efectuar, uma para x e outra para y . Isto significa que em vez de uma função g que aparece na equação (11.29), temos agora duas funções, a saber X e Y , as quais relacionam x, y com u, v do modo seguinte:

$$(11.30) \quad x = X(u, v), \quad y = Y(u, v).$$

As duas equações (11.30) definem uma aplicação que faz corresponder a um ponto (u, v) do plano UOV , o ponto (x, y) do plano XOY . Um conjunto T de pontos no plano UOV é aplicado sobre outro conjunto S do plano XOY , como se sugere na figura 11.23. A aplicação pode também exprimir-se mediante uma função vectorial. No plano XOY , e a partir da origem, traçamos o raio vector r para um ponto genérico (x, y) de S , como se indica na figura 11.23. O vector r depende simultaneamente de u e v e pode considerar-se como uma função vectorial de duas variáveis definida pela equação

$$(11.31) \quad r(u, v) = X(u, v)\mathbf{i} + Y(u, v)\mathbf{j} \quad \text{se } (u, v) \in T.$$

Esta equação chama-se a *equação vectorial* da aplicação. Quando (u, v) percorre a região T , a extremidade de $r(u, v)$ descreve pontos de S .

Algumas vezes as duas equações em (11.30) podem resolver-se de modo a darem u e v em função de x e y . Quanto tal é possível podemos exprimir o resultado na forma

$$u = U(x, y), \quad v = V(x, y).$$

Estas equações definem uma aplicação do plano XOY no plano UOV , chamada *aplicação inversa* da definida por (11.30), uma vez que transforma pontos de S em pontos de T . As chamadas *aplicações biunívocas* são de importância fundamental. Estas transformam pontos *distintos* de T em pontos *distintos* de S ; por outras palavras, dois pontos distintos de T não podem transformar-se no mesmo ponto de S mediante uma aplicação biunívoca. Tais aplicações estabelecem uma correspondência biunívoca entre os pontos de T e os correspondentes de S e permitem-nos (pelo menos teoricamente) voltar de S a T pela aplicação inversa (que, naturalmente, é também biunívoca).

Consideraremos aplicações para as quais as funções X e Y são contínuas e têm derivadas parciais contínuas $\partial X/\partial u$, $\partial X/\partial v$, $\partial Y/\partial u$ e $\partial Y/\partial v$ em S . Hipóteses análogas se admitem para as funções U e V . Estas hipóteses não constituem restrições muito fortes visto serem satisfeitas pela maior parte das funções que surgem na prática.

A fórmula para mudança de variáveis nos integrais duplos pode escrever-se:

$$(11.32) \quad \iint_S f(x, y) \, dx \, dy = \iint_T f[X(u, v), Y(u, v)] |J(u, v)| \, du \, dv.$$

O factor $J(u, v)$, que aparece no integrando do segundo membro, desempenha o papel do factor $g(t)$ que figura na fórmula (11.29). Este factor chama-se o *determinante Jacobiano* da aplicação definida por (11.30); é igual a

$$J(u, v) = \begin{vmatrix} \frac{\partial X}{\partial u} & \frac{\partial X}{\partial v} \\ \frac{\partial Y}{\partial u} & \frac{\partial Y}{\partial v} \end{vmatrix}$$

Algumas vezes usa-se o símbolo $\partial(X, Y)/\partial(u, v)$, em vez de $J(u, v)$, para representar o jacobiano.

Não discutiremos as condições mais gerais sob as quais é válida a fórmula de transformação (11.32). Pode provar-se† que (11.32) é válida se, em complemento das hipóteses de continuidade de X , Y , U e V já referidas, supomos que a aplicação de T em S é biunívoca e que o jacobiano $J(u, v)$ nunca se anula. A fórmula é ainda válida se a aplicação não for biunívoca apenas num subconjunto de T de medida nula, ou se o Jacobiano se anula num subconjunto de medida nula.

Na Secção 11.30 mostraremos como a fórmula de transformação (11.32) pode deduzir-se como uma consequência de um dos seus casos particulares, concretamente, o caso em que S é um rectângulo e a função f tem o valor constante 1 em cada ponto de S . Neste caso particular (11.32) vem

$$(11.33) \quad \iint_S dx \, dy = \iint_T |J(u, v)| \, du \, dv.$$

Ainda para este caso a demonstração não é simples. Na Secção 11.29 é dada uma demonstração de (11.33) com recurso ao teorema de Green. A parte restante desta Secção é destinada à apresentação de um argumento geométrico simples o qual justifica a validade da fórmula (11.33).

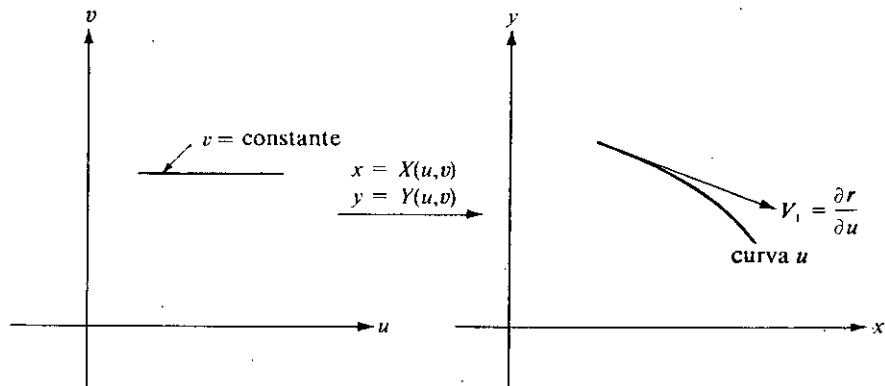


FIG. 11.24. Uma curva u e o correspondente vector velocidade.

Interpretação geométrica da fórmula (11.33). Tomemos uma região T no plano UOV , como se indica na figura 11.23, e seja S o conjunto dos pontos do plano XOY sobre os quais é aplicado T , por intermédio da função vectorial r dada por (11.31). Introduzamos agora duas novas funções vectoriais V_1 e V_2 que se obtêm tomando as derivadas parciais dos componentes de r com respeito a u e v , respectivamente.

† Ver teorema 10.30 do livro do mesmo autor *Mathematical Analysis* (trad. espanhola "Análisis Matemático").

Quer dizer, definimos

$$\mathbf{V}_1 = \frac{\partial \mathbf{r}}{\partial u} = \frac{\partial X}{\partial u} \mathbf{i} + \frac{\partial Y}{\partial u} \mathbf{j} \quad \text{e} \quad \mathbf{V}_2 = \frac{\partial \mathbf{r}}{\partial v} = \frac{\partial X}{\partial v} \mathbf{i} + \frac{\partial Y}{\partial v} \mathbf{j}.$$

Estes vectores podem interpretar-se geometricamente do modo seguinte: Consideremos um segmento de recta no plano UOV e paralelo a OU (v é constante ao longo de tal segmento): A função vectorial \mathbf{r} aplica este segmento sobre uma curva (chamada a curva u) no plano XOY , como sugere a figura 11.24. Se considerarmos o parâmetro u como representando o «tempo», o vector \mathbf{V}_1 representa a velocidade do vector posicional \mathbf{r} e é portanto tangente à curva traçada pela extremidade de \mathbf{r} . Do mesmo modo, cada vector \mathbf{V}_2 representa o vector velocidade de uma curva v obtida fazendo-se $u = \text{constante}$. Por cada ponto da região S passam uma curva u e uma curva v .

Consideremos agora um pequeno rectângulo com lados Δu e Δv , como se indica na figura 11.25. Se Δu é a medida de um pequeno intervalo de tempo então um ponto da curva u , durante o tempo Δu , desloca-se ao longo da curva aproximadamente do valor do produto $\|\mathbf{V}_1\| \Delta u$ (visto que $\|\mathbf{V}_1\|$ representa a velocidade e Δu o intervalo de tempo). Analogamente, durante o tempo Δv um ponto da curva v move-se de uma distância aproximadamente igual a $\|\mathbf{V}_2\| \Delta v$. Por isso a região rectangular com lados

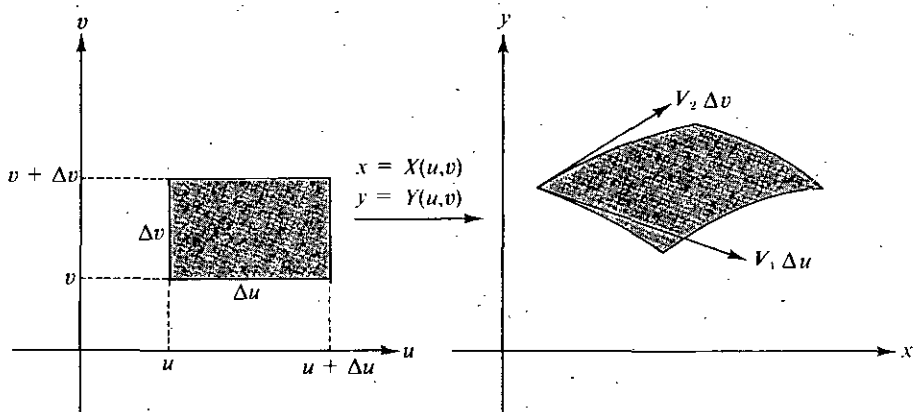


FIG. 11.25. A imagem de uma região rectangular no plano UOV é um paralelogramo curvo no plano XOY .

Δu e Δv no plano UOV transforma-se numa porção do plano XOY que é um quase paralelogramo, cujos lados são os vectores $\mathbf{V}_1 \Delta u$ e $\mathbf{V}_2 \Delta v$ como se vê na figura 11.25. A área deste paralelogramo é dada pela grandeza do produto vectorial dos vectores $\mathbf{V}_1 \Delta u$ e $\mathbf{V}_2 \Delta v$; esta é igual a

$$\|(\mathbf{V}_1 \Delta u) \times (\mathbf{V}_2 \Delta v)\| = \|\mathbf{V}_1 \times \mathbf{V}_2\| \Delta u \Delta v.$$

Se calculamos o produto $\mathbf{V}_1 \times \mathbf{V}_2$ em função das componentes dos vectores \mathbf{V}_1 e \mathbf{V}_2 vem

$$V_1 \times V_2 = \begin{vmatrix} i & j & k \\ \frac{\partial X}{\partial u} & \frac{\partial Y}{\partial u} & 0 \\ \frac{\partial X}{\partial v} & \frac{\partial Y}{\partial v} & 0 \end{vmatrix} = \begin{vmatrix} \frac{\partial X}{\partial u} & \frac{\partial Y}{\partial u} \\ \frac{\partial X}{\partial v} & \frac{\partial Y}{\partial v} \end{vmatrix} k = J(u, v)k.$$

Portanto a grandeza de $V_1 \times V_2$ é exactamente $|J(u, v)|$ e a área do paralelogramo curvo da figura 11.25 é quase igual a $|J(u, v)| \Delta u \Delta v$.

Se $J(u, v) = 1$ para todos os pontos de T , o "paralelogramo" tem a mesma área que o rectângulo e a aplicação conserva a área. Caso contrário, para obtermos a área do paralelogramo temos que multiplicar a área do rectângulo por $|J(u, v)|$. Isto sugere que o jacobiano pode considerar-se como um "factor de ampliação" das áreas.

Seja P uma partição de um rectângulo R , suficientemente grande para conter toda a região T e consideremos um subrectângulo genérico de P de, digamos, lados Δu e Δv . Se Δu e Δv são pequenos, o jacobiano J é quase constante nesse rectângulo e consequentemente J actua de modo algo semelhante a uma função em escada em R . (Atribuímos a J o valor zero no exterior de T .) Se consideramos J como uma autêntica função em escada, então o integral duplo de $|J|$ sobre R (e consequentemente sobre T) é uma soma de produtos da forma $|J(u, v)| \Delta u \Delta v$ e as observações efectuadas sugerem que esta soma é aproximadamente igual à área de S , que sabemos ser o valor do integral duplo $\int_S dx dy$.

Esta discussão geométrica, que apenas sugere porquê podemos esperar a validade de uma fórmula como (11.33), pode tornar-se a base de uma demonstração rigorosa, mas os pormenores são muito extensos e, mais ainda, complicados. Como referimos atrás, uma demonstração de (11.33), seguindo uma via totalmente diferente, será dada mais adiante.

Se $J(u, v) = 0$ num ponto particular (u, v) , os dois vectores V_1 e V_2 são paralelos (visto ser nulo o respectivo produto vectorial) e o paralelogramo degenera num segmento de recta. Tais pontos dizem-se *pontos singulares* da aplicação. Como já referimos, a fórmula de transformação (11.32) é também válida se existir apenas um número finito de tais pontos singulares ou, mais geralmente, quando os pontos singulares formam um conjunto de medida nula. Isto é o que se verifica em todas as aplicações que usaremos. Na próxima secção ilustramos o uso da fórmula (11.32) com dois exemplos importantes.

11.27. Casos particulares da fórmula de mudança de variáveis

EXEMPLO 1. Coordenadas polares. Neste caso escrevemos r e θ em vez de u e v e definimos a aplicação pelas duas equações:

$$x = r \cos \theta, \quad y = r \sin \theta.$$

Quer dizer, $X(r, \theta) = r \cos \theta$, $Y(r, \theta) = r \sin \theta$. Para obtermos uma aplicação biunívoca fazemos $r > 0$ e restringimos θ aos valores de um intervalo da forma $\theta_0 \leq \theta \leq \theta_0 + 2\pi$.

Por exemplo, a aplicação é biunívoca em qualquer subconjunto do rectângulo $(0, a] \times [0, 2\pi)$ no plano $r\theta$. O determinante jacobiano desta aplicação é

$$J(r, \theta) = \begin{vmatrix} \frac{\partial X}{\partial r} & \frac{\partial Y}{\partial r} \\ \frac{\partial X}{\partial \theta} & \frac{\partial Y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos \theta & \sin \theta \\ -r \sin \theta & r \cos \theta \end{vmatrix} = r(\cos^2 \theta + \sin^2 \theta) = r.$$

Logo a fórmula (11.32) vem

$$\iint_S f(x, y) \, dx \, dy = \iint_T f(r \cos \theta, r \sin \theta) \, r \, dr \, d\theta.$$

As curvas r são rectas passando pela origem e as curvas θ são círculos centrados na origem. A imagem de um rectângulo no plano $Or\theta$ é um “paralelogramo” no plano XOY limitado por raios e dois arcos de circunferência, como se mostra na figura 11.26

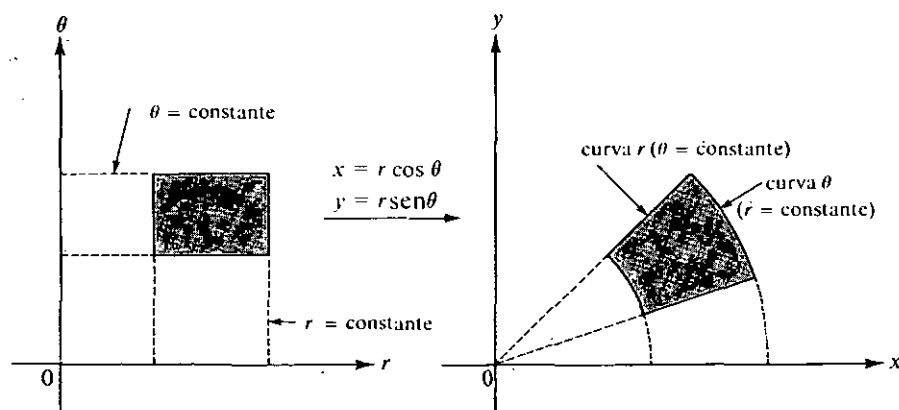


FIG. 11.26. Transformação por coordenadas polares.

O jacobiano anula-se quando $r = 0$, mas tal facto não afecta a validade da fórmula de transformação porque o conjunto de pontos com $r = 0$ tem medida nula.

Visto que $V_1 = \cos \theta \, i + \sin \theta \, j$, temos $\|V_1\| = 1$, pelo que não há distorção das distâncias ao longo das curvas r . Por outro lado, temos

$$V_2 = -r \sin \theta \, i + r \cos \theta \, j, \quad \|V_2\| = r,$$

pelo que as distâncias ao longo da curva θ vêm multiplicadas pelo factor r .

As coordenadas polares são particularmente bem adequadas quando a região de integração tem fronteiras ao longo das quais r ou θ são constantes. Por exemplo, consideremos o integral que permite o cálculo do volume de um octante de esfera de raio a ,

$$\iint_S \sqrt{a^2 - x^2 - y^2} \, dx \, dy,$$

onde a região S é o primeiro quadrante do disco circular $x^2 + y^2 \leq a^2$. Em coordenadas polares o volume vem

$$\iint_T \sqrt{a^2 - r^2} \, r \, dr \, d\theta,$$

em que a região de integração T é agora um rectângulo $[\theta, a] \times [0, \frac{1}{2}\pi]$. Integrando primeiramente em relação a θ e em seguida em relação a r obtemos

$$\iint_T \sqrt{a^2 - r^2} \, r \, dr \, d\theta = \frac{\pi}{2} \int_0^a r \sqrt{a^2 - r^2} \, dr = \frac{\pi}{2} \cdot \frac{(a^2 - r^2)^{3/2}}{-3} \Big|_0^a = \frac{\pi a^3}{6}.$$

O mesmo resultado pode ser obtido considerando coordenadas rectangulares, mas os cálculos são mais complicados.

EXEMPLO 2. Transformações lineares. Uma transformação linear é uma aplicação definida por um par de equações da forma

$$(11.34) \quad x = Au + Bv, \quad y = Cu + Dv,$$

onde A, B, C, D são constantes dadas. O determinante jacobiano é

$$J(u, v) = AD - BC,$$

e para assegurar a transformação inversa supomos que $AD - BC \neq 0$. Isto garante-nos que as duas equações lineares (11.34) podem resolver-se de modo a obtermos u e v em função de x e y .

As transformações lineares transformam rectas paralelas em rectas paralelas. Portanto a imagem de um rectângulo no plano uv é um paralelogramo no plano XOY , e a sua área é a do rectângulo multiplicada pelo factor $|J(u, v)| = |AD - BC|$. A fórmula (11.32) vem então

$$\iint_S f(x, y) \, dx \, dy = |AD - BC| \iint_T f(Au + Bv, Cu + Dv) \, du \, dv.$$

Para ilustrar um exemplo em que é útil uma mudança de variáveis linear consideremos o integral

$$\iint_S e^{(y-x)/(y+x)} dx dy,$$

onde S é o triângulo limitado pela recta $x + y = 2$ e pelos dois eixos coordenados. (Ver figura 11.27). A presença de $y - x$ e $y + x$ no integrando sugere a mudança de variáveis

$$u = y - x, \quad v = y + x.$$

Resolvendo relativamente a x e a y encontramos

$$x = \frac{v - u}{2} \quad \text{e} \quad y = \frac{v + u}{2}.$$

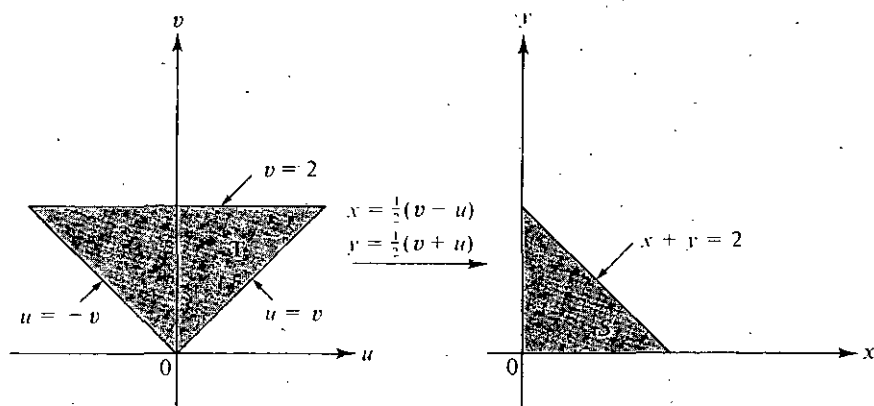


FIG. 11.27. Aplicação por uma transformação linear.

O determinante jacobiano é $J(u, v) = -\frac{1}{2}$. Para determinar a imagem T de S no plano uv observamos que as rectas $x = 0$ e $y = 0$ aplicam-se nas rectas $u = v$ e $u = -v$, respectivamente; a recta $x + y = 2$ transforma-se na recta $v = 2$. Pontos interiores a S verificam $0 < x + y < 2$ e transformam-se em pontos de T verificando $0 < v < 2$. Portanto a nova região de integração T é uma região triangular, como se indica na figura 11.27. O integral duplo em questão vem

$$\iint_S e^{(y-x)/(y+x)} dx dy = \frac{1}{2} \iint_T e^{u/v} du dv.$$

Integrando em primeiro lugar em relação a u encontramos

$$\frac{1}{2} \iint_T e^{u/v} du dv = \frac{1}{2} \int_0^2 \left[\int_{-v}^v e^{u/v} du \right] dv = \frac{1}{2} \int_0^2 v \left(e - \frac{1}{e} \right) dv = e - \frac{1}{e}.$$

11.28. Exercícios

Em cada um dos Exercícios 1 a 5, fazer um desenho da região S e exprimir o integral duplo $\iint_S f(x, y) dx dy$ por integrais repetidos em coordenadas polares

1. $S = \{(x, y) \mid x^2 + y^2 \leq a^2\}$, onde $a > 0$.
2. $S = \{(x, y) \mid x^2 + y^2 \leq 2x\}$.
3. $S = \{(x, y) \mid a^2 \leq x^2 + y^2 \leq b^2\}$, onde $0 < a < b$.
4. $S = \{(x, y) \mid 0 \leq y \leq 1 - x, 0 \leq x \leq 1\}$.
5. $S = \{(x, y) \mid x^2 \leq y \leq 1, -1 \leq x \leq 1\}$.

Em cada um dos Exercícios 6 a 9, mudar o integral para coordenadas polares e calcular o seu valor. (A letra a representa uma constante positiva.)

6. $\int_0^{2a} \left[\int_0^{\sqrt{2ax-x^2}} (x^2 + y^2) dy \right] dx$.
8. $\int_0^1 \left[\int_{x^2}^x (x^2 + y^2)^{-1/2} dy \right] dx$.
7. $\int_0^a \left[\int_0^x \sqrt{x^2 + y^2} dy \right] dx$.
9. $\int_0^a \left[\int_0^{\sqrt{a^2-y^2}} (x^2 + y^2) dx \right] dy$.

Nos exercícios 10 a 13, transformar cada um dos integrais dados em um ou mais integrais repetidos em coordenadas polares

10. $\int_0^1 \left[\int_0^1 f(x, y) dy \right] dx$.
12. $\int_0^1 \left[\int_{1-x}^{\sqrt{1-x^2}} f(x, y) dy \right] dx$.
11. $\int_0^2 \left[\int_x^{x\sqrt{3}} f(\sqrt{x^2 + y^2}) dy \right] dx$.
13. $\int_0^1 \left[\int_0^{x^2} f(x, y) dy \right] dx$.

14. Recorrer a uma transformação linear conveniente para calcular o integral duplo

$$\iint_S (x - y)^2 \sin^2(x + y) dx dy$$

sendo S a paralelogramo de vértices $(\pi, 0)$, $(2\pi, \pi)$, $(\pi, 2\pi)$, $(0, \pi)$.

15. Um paralelogramo S do plano XOY tem por vértices $(0, 0)$, $(2, 10)$, $(3, 17)$, e $(1, 7)$.
 - (a) Determinar uma transformação linear $u = ax + by$, $v = cx + dy$, que aplique S num retângulo R do plano uv com vértices opostos $(0, 0)$ e $(4, 2)$. O vértice $(2, 10)$ deverá aplicar-se num ponto do eixo u .
 - (b) Calcular o integral duplo $\iint_S xy dx dy$ transformando-o num integral equivalente sobre o rectângulo R da alínea (a).
16. Se $r > 0$, seja $I(r) = \int_R e^{-u^2} du$.
 - (a) Mostrar que $I^2(r) = \iint_R e^{-(x^2 + y^2)} dx dy$, sendo R o quadrado $R = [-r, r] \times [-r, r]$.
 - (b) Se C_1 e C_2 são discos circulares inscrevendo e circunscrevendo R , provar que

$$\iint_{C_1} e^{-(x^2 + y^2)} dx dy < I^2(r) < \iint_{C_2} e^{-(x^2 + y^2)} dx dy$$
 - (c) Expressar os integrais estendidos a C_1 e C_2 em coordenadas polares e usar (b) para deduzir que $I(r) \rightarrow \sqrt{\pi}$ quando $r \rightarrow \infty$. Isto demonstra que $\int_0^\infty e^{-u^2} du = \sqrt{\pi}/2$.
 - (d) Usar a alínea (c) para provar que $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, com Γ o símbolo da função gama.

17. Considerar a aplicação definida pelas equações

$$x = u + v, \quad y = v - u^2.$$

- (a) Calcular o jacobiano $J(u, v)$.
 (b) Um triângulo T no plano uv tem vértices $(0, 0)$, $(2, 0)$, $(0, 2)$. Deesenhar a sua imagem S no plano XOY .
 (c) Calcular a área de S por intermédio de um integral duplo estendido a S e também por intermédio de um integral duplo estendido a T .
 (d) Calcular $\iint_S (x - y + 1)^{-2} dx dy$.

18. Considerar a aplicação definida pelas duas equações $x = u^2 - v^2$, $y = 2uv$.

- (a) Calcular o determinante jacobiano $J(u, v)$.
 (b) Designe-se por T o rectângulo no plano uv com vértices $(1, 1)$, $(2, 1)$, $(2, 3)$, $(1, 3)$. Representar, por meio de um desenho, a imagem S no plano XOY .
 (c) Calcular o integral duplo $\iint_C xy dx dy$ recorrendo à mudança de variáveis $x = u^2 - v^2$, com $C = \{(x, y) \mid x^2 + y^2 \leq 1\}$.

19. Calcular o integral duplo

$$I(p, r) = \iint_R \frac{dx dy}{(p^2 + x^2 + y^2)^p}$$

sobre o disco circular $R = \{(x, y) \mid x^2 + y^2 \leq r^2\}$. Determinar os valores de p para os quais $I(p, r)$ tende para um limite quando $r \rightarrow +\infty$.

Nos Exercícios 20 a 22 provar as igualdades dadas por introdução de uma adequada mudança de variáveis em cada caso.

$$20. \iint_S f(x+y) dx dy = \int_{-1}^1 f(u) du, \quad \text{onde } S = \{(x, y) \mid |x| + |y| \leq 1\}.$$

$$21. \iint_S f(ax + by + c) dx dy = 2 \int_{-1}^1 \sqrt{1-u^2} f(u\sqrt{a^2+b^2} + c) du,$$

onde $S = \{(x, y) \mid x^2 + y^2 \leq 1\}$ e $a^2 + b^2 \neq 0$.

$$22. \iint_S f(xy) dx dy = \log 2 \int_1^2 f(u) du, \quad \text{onde } S \text{ é a região do primeiro quadrante limitada pelas curvas } xy = 1, xy = 2, x = y, y = 4x.$$

11.29. Demonstração da fórmula de mudança de variáveis num caso particular

Como já referimos, a fórmula

$$(11.35) \quad \iint_S f(x, y) dx dy = \iint_T f[X(u, v), Y(u, v)] |J(u, v)| du dv$$

pode deduzir-se como uma consequência do caso particular em que S é um rectângulo e f é identicamente 1. Neste caso a fórmula simplifica-se para

$$(11.36) \quad \iint_R dx \, dy = \iint_{R^*} |J(u, v)| \, du \, dv.$$

Aqui R representa um rectângulo no plano XOY e R^* a sua imagem no plano UV . (Ver figura 11.28) obtida por uma aplicação biunívoca

$$u = U(x, y), \quad v = V(x, y).$$

A aplicação inversa é dada por

$$x = X(u, v), \quad y = Y(u, v),$$

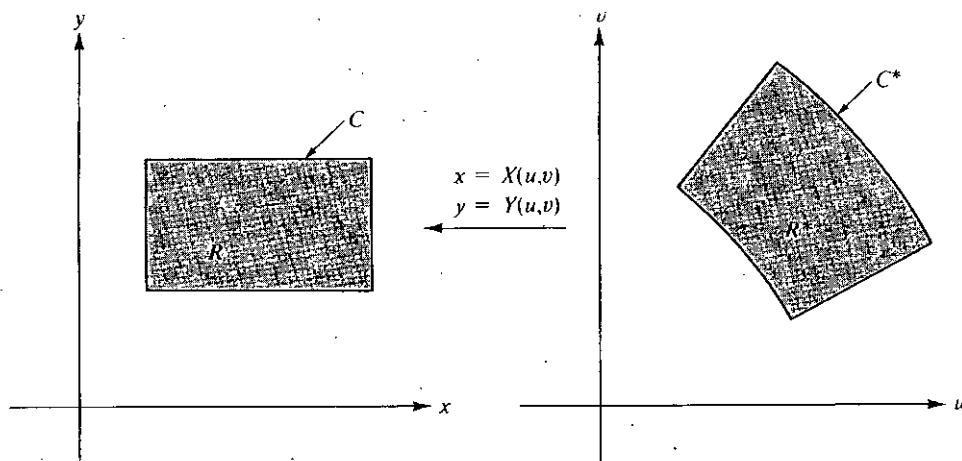


FIG. 11.28. A fórmula de mudança de variáveis para integrais duplos deduzida do teorema de Green.

e $J(u, v)$ representa o determinante jacobiano,

$$J(u, v) = \begin{vmatrix} \frac{\partial X}{\partial u} & \frac{\partial X}{\partial v} \\ \frac{\partial Y}{\partial u} & \frac{\partial Y}{\partial v} \end{vmatrix}.$$

Nesta Secção usamos o teorema de Green para provarmos (11.36), e na Secção seguinte deduzimos a fórmula mais geral (11.35) a partir do caso particular (11.36).

Para a demonstração supomos que as funções X e Y possuem derivadas parciais contínuas de segunda ordem e que o jacobiano nunca se anula em R^* . Então $J(u, v)$ é sempre positivo, ou sempre negativo. O significado do sinal de $J(u, v)$ é o seguinte:

quando um ponto (x, y) descreve a fronteira de R no sentido directo, o ponto imagem (u, v) descreve o contorno de R^* no sentido directo se $J(u, v)$ é positivo e no sentido inverso se $J(u, v)$ é negativo. Na demonstração vamos supor que $J(u, v) > 0$.

A ideia da demonstração consiste em exprimir cada integral duplo de (11.36) como um integral de linha, recorrendo ao teorema de Green. Verificamos depois a igualdade dos dois integrais de linha exprimindo cada um deles na forma paramétrica.

Comencemos com o integral duplo no plano XOY , escrevendo

$$\iint_R dx dy = \iint_R \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx dy,$$

onde $Q(x, y) = x$ e $P(x, y) = 0$. Pelo teorema de Green este integral duplo é igual ao integral de linha

$$\int_C P dx + Q dy = \int_C x dy.$$

Aqui C é a fronteira de R , descrita no sentido directo. Analogamente, transformamos o integral duplo do plano UV num integral de linha ao longo da fronteira C^* de R^* . O integrando, $J(u, v)$, pode escrever-se

$$\begin{aligned} J(u, v) &= \frac{\partial X}{\partial u} \frac{\partial Y}{\partial v} - \frac{\partial X}{\partial v} \frac{\partial Y}{\partial u} = \frac{\partial X}{\partial u} \frac{\partial Y}{\partial v} + X \frac{\partial^2 Y}{\partial u \partial v} - X \frac{\partial^2 Y}{\partial v \partial u} - \frac{\partial X}{\partial v} \frac{\partial Y}{\partial u} \\ &= \frac{\partial}{\partial u} \left(X \frac{\partial Y}{\partial v} \right) - \frac{\partial}{\partial v} \left(X \frac{\partial Y}{\partial u} \right). \end{aligned}$$

Aplicando o teorema de Green ao integral duplo sobre R^* encontramos

$$\iint_{R^*} J(u, v) du dv = \int_{C^*} \left(X \frac{\partial Y}{\partial u} du + X \frac{\partial Y}{\partial v} dv \right).$$

Deste modo, para completarmos a demonstração de (11.36) necessitamos somente provar que

$$(11.37) \quad \int_C x dy = \int_{C^*} \left(X \frac{\partial Y}{\partial u} du + X \frac{\partial Y}{\partial v} dv \right).$$

Introduzimos uma parametrização para C^* e utilizemo-la para achar uma representação de C . Suponhamos que C^* é definida por uma função α num intervalo $[a, b]$, seja

$$\alpha(t) = U(t)i + V(t)j.$$

Designemos por

$$\beta(t) = X[U(t), V(t)]i + Y[U(t), V(t)]j.$$

Então quando t varia no intervalo $[a, b]$, o vector $\alpha(t)$ descreve a curva C^* e $\beta(t)$ a curva C . A derivada de β é dada por

$$\beta'(t) = \left[\frac{\partial X}{\partial u} U'(t) + \frac{\partial X}{\partial v} V'(t) \right] i + \left[\frac{\partial Y}{\partial u} U'(t) + \frac{\partial Y}{\partial v} V'(t) \right] j.$$

Logo

$$\int_C x \, dy = \int_a^b X[U(t), V(t)] \left(\frac{\partial Y}{\partial u} U'(t) + \frac{\partial Y}{\partial v} V'(t) \right) dt.$$

O último integral estendido a $[a, b]$ obtêm-se igualmente parametrando o integral de linha sobre C^* em (11.37). Portanto os dois integrais de linha em (11.37) são iguais, o que prova (11.36).

11.30. Demonstração da fórmula de mudança de variáveis no caso geral

Nesta Secção deduzimos a fórmula geral de mudança de variáveis,

$$(11.38) \quad \iint_S f(x, y) \, dx \, dy = \iint_T f[X(u, v), Y(u, v)] |J(u, v)| \, du \, dv$$

a partir do caso particular tratado na Secção anterior,

$$(11.39) \quad \iint_R dx \, dy = \iint_{R^*} |J(u, v)| \, du \, dv,$$

em que R é um rectângulo e R^* a sua imagem no plano uv .

Demonstramos primeiramente que

$$(11.40) \quad \iint_R s(x, y) \, dx \, dy = \iint_{R^*} s[X(u, v), Y(u, v)] |J(u, v)| \, du \, dv,$$

onde s é qualquer função em escada definida em R . Com esta finalidade, seja P uma partição de R em mn subrectângulos R_{ij} de lados Δx_i e Δy_j , e seja c_{ij} o valor constante que s toma no subrectângulo aberto R_{ij} . Aplicando (11.39) ao rectângulo R_{ij} encontramos

$$\Delta x_i \Delta y_j = \iint_{R_{ij}} dx \, dy = \iint_{R_{ij}^*} |J(u, v)| \, du \, dv.$$

Multiplicando ambos os membros por c_{ij} e somando a respeito dos índices i e j obtemos

$$(11.41) \quad \sum_{i=1}^n \sum_{j=1}^m c_{ij} \Delta x_i \Delta y_j = \sum_{i=1}^n \sum_{j=1}^m c_{ij} \iint_{R_{ij}^*} |J(u, v)| \, du \, dv.$$

Uma vez que s é uma função em escada, a igualdade anterior é equivalente a

$$(11.42) \quad \sum_{i=1}^n \sum_{j=1}^m c_{ij} \Delta x_i \Delta y_j = \sum_{i=1}^n \sum_{j=1}^m \iint_{R_{ij}^*} s[X(u, v), Y(u, v)] |J(u, v)| du dv.$$

Recorrendo à propriedade aditiva dos integrais duplos vemos que (11.42) é a mesma que (11.40). Então, (11.40) é uma consequência de (11.39).

Seguidamente vamos demonstrar que a função em escada s , em (11.40), pode ser substituída por qualquer função f para a qual existam ambos os membros de (11.40). Seja f integrável sobre um rectângulo R e escolhamos funções em escada s e t verificando as desigualdades

$$(11.43) \quad s(x, y) \leq f(x, y) \leq t(x, y),$$

para todos os pontos (x, y) de R . Então verifica-se também

$$(11.44) \quad s[X(u, v), Y(u, v)] \leq f[X(u, v), Y(u, v)] \leq t[X(u, v), Y(u, v)]$$

para todo o ponto (u, v) pertencente à imagem R^* . Por uma questão de brevidade, escrevemos $S(u, v)$ em lugar de $s[X(u, v), Y(u, v)]$ e definimos $F(u, v)$ e $T(u, v)$ do mesmo modo. Multiplicando as desigualdades (11.44) por $|J(u, v)|$ e integrando sobre R^* obtemos

$$\iint_{R^*} S(u, v) |J(u, v)| du dv \leq \iint_{R^*} F(u, v) |J(u, v)| du dv \leq \iint_{R^*} T(u, v) |J(u, v)| du dv.$$

Devido a (11.40), as desigualdades precedentes são as mesmas que

$$\iint_R s(x, y) dx dy \leq \iint_{R^*} F(u, v) |J(u, v)| du dv \leq \iint_R t(x, y) dx dy.$$

Portanto $\iint_{R^*} F(u, v) |J(u, v)| du dv$ é um número compreendido entre os integrais $\iint_R s(x, y) dx dy$ e $\iint_R t(x, y) dx dy$ para qualquer par de funções R em escada s e t verificando (11.43). Por f ser integrável, implica que

$$\iint_R f(x, y) dx dy = \iint_{R^*} F(u, v) |J(u, v)| du dv$$

e por conseguinte (11.38) é válida para funções integráveis definidas sobre rectângulos.

Uma vez demonstrada a validade de (11.38) para rectângulos, podemos facilmente generalizá-la a regiões mais gerais S pelo processo habitual de definir um rectângulo R contendo S e substituindo a função f por uma outra, \tilde{f} , que coincida com f em S e se anule no exterior de S . Então temos que

$$\iint_S f = \iint_R \tilde{f} = \iint_{R^*} \tilde{f}[X(u, v), Y(u, v)] |J(u, v)| du dv = \iint_T F(u, v) |J(u, v)| du dv$$

o que demonstra que (11.38) é, na verdade, uma consequência de (11.39).

11.31. Extensões a um número superior de dimensões

O conceito de integral múltiplo pode generalizar-se do espaço bidimensional para um espaço a n dimensões, com n qualquer ≥ 3 . Visto que o desenvolvimento do assunto é completamente análogo ao caso em que $n = 2$, referimos apenas os resultados principais.

O integrando é um campo escalar f definido e limitado num conjunto S do espaço n dimensional. O integral de f sobre S , chamado um integral n -múltiplo, representa-se pelo símbolo

$$\int_S \cdots \int_S f, \quad \text{ou} \quad \int_S \cdots \int_S f(x_1, \dots, x_n) dx_1 \cdots dx_n,$$

com n sinais de integral, ou mais simplesmente com um sinal de integral, $\int_S f(x) dx$, onde $x = (x_1, \dots, x_n)$. Quando $n = 3$ escrevemos (x, y, z) em vez de (x_1, x_2, x_3) e representamos os integrais tripos por

$$\iiint_S f, \quad \text{ou} \quad \iiint_S f(x, y, z) dx dy dz.$$

Em primeiro lugar definimos o integral n -múltiplo para uma função em escada definida num intervalo n -dimensional. Lembramos que um intervalo n -dimensional fechado $[a, b]$ é o produto cartesiano de n intervalos unidimensionais fechados $[a_k, b_k]$, com $a = (a_1, \dots, a_n)$ e $b = (b_1, \dots, b_n)$. Um intervalo aberto n -dimensional (a, b) é o produto cartesiano de n intervalos abertos (a_k, b_k) . O volume de $[a, b]$, ou de (a, b) , define-se como o produto dos comprimentos dos intervalos componentes

$$(b_1 - a_1) \cdots (b_n - a_n).$$

Se P_1, \dots, P_n são partições de $[a_1, b_1], \dots, [a_n, b_n]$, respectivamente, o produto cartesiano $P = P_1 \times \dots \times P_n$ diz-se uma partição de $[a, b]$. Uma função f , definida em $[a, b]$, diz-se uma função em escada se é constante sobre cada um dos subintervalos abertos definidos por alguma partição P . O integral de ordem n de uma tal função em escada define-se pela fórmula

$$\int_{[a, b]} \cdots \int_{[a, b]} f = \sum_i c_i v_i,$$

onde c_i é o valor constante que f toma no subintervalo de ordem i e v_i é o seu volume. A soma é uma soma finita estendida a todos os subintervalos de P .

A partir da definição de integral n -múltiplo para funções em escada, definimos o integral para funções limitadas mais gerais definidas em intervalos, seguindo o processo usual. Sejam s e t funções em escada tais que $s \leq f \leq t$ em $[a, b]$. Se existir um ϵ e um só número I tal que

$$\int \cdots \int_{[a,b]} s \leq I \leq \int \cdots \int_{[a,b]} t$$

quaisquer que sejam s e t satisfazendo a $s \leq f \leq t$, então f diz-se ser integrável em $[a, b]$ e o número I chama-se o integral n -múltiplo de f ,

$$I = \int \cdots \int_{[a,b]} f.$$

Tal como no caso bidimensional, o integral existe se f é contínua em $[a, b]$. Também existe se f é limitada em $[a, b]$ e se o conjunto de descontinuidades de f tem medida n -dimensional nula. Um conjunto limitado S tem medida n -dimensional nula se para cada $\epsilon > 0$ existe um conjunto finito de intervalos n -dimensionais cuja união inclui S e tal que soma dos respectivos volumes não excede ϵ .

Para definirmos um integral n -múltiplo de uma função limitada f sobre um conjunto limitado S mais geral, substituímos f por uma nova função \tilde{f} , a qual coincide com f em S e se anula fora de S ; o integral de f sobre S define-se como o integral de \tilde{f} sobre algum intervalo que contenha S .

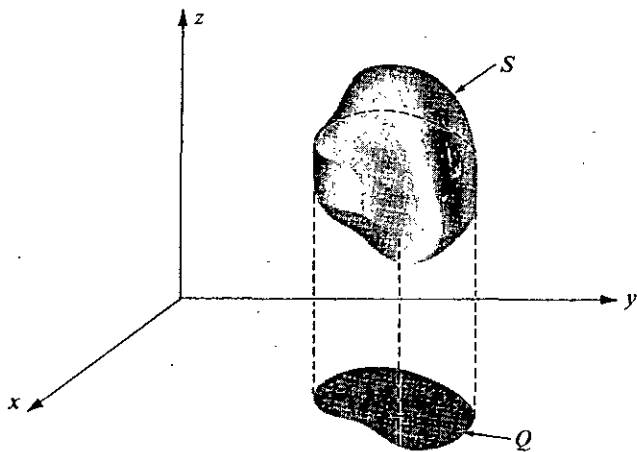
Alguns integrais múltiplos podem calcular-se pelo uso repetido de integrais de dimensão inferior. Por exemplo, suponhamos que S é um conjunto no espaço tridimensional definido do modo seguinte:

$$(11.45) \quad S = \{(x, y, z) \mid (x, y) \in Q \quad \text{e} \quad \varphi_1(x, y) \leq z \leq \varphi_2(x, y)\},$$

onde Q é uma região bidimensional, chamada a projecção de S no plano XOY , e φ_1, φ_2 são contínuas em S . (Na figura 11.29 mostra-se um exemplo). Conjuntos deste tipo são limitados por duas superfícies com equações cartesianas $z = \varphi_1(x, y)$, e $z = \varphi_2(x, y)$ e (talvez) uma porção do cilindro gerado por uma recta movendo-se paralelamente ao eixo OZ ao longo do contorno de Q . As rectas paralelas ao eixo OZ intersectam este conjunto segundo segmentos de rectas que unem a superfície inferior com a superfície superior. Se f é contínua no interior de S , temos a formula para as integrações repetidas

$$(11.46) \quad \iiint_S f(x, y, z) \, dx \, dy \, dz = \iint_Q \left[\int_{\varphi_1(x,y)}^{\varphi_2(x,y)} f(x, y, z) \, dz \right] dx \, dy.$$

Quer dizer, para x e y fixos, a primeira integração é efectuada relativamente a z desde a superfície fronteira inferior até à superior. Isto reduz o cálculo a um integral duplo sobre a projecção Q , o qual pode ser resolvido pelos métodos já expostos atrás.

FIG. 11.29. Um sólido S e a sua projecção Q no plano XOY .

Há mais dois tipos de conjuntos análogos aos descritos em (11.45) para os quais os eixos OX e OY desempenham o papel do eixo OZ , sendo as projecções tomadas respectivamente sobre YOZ e XOZ . Os integrais triplos sobre tais conjuntos podem calcular-se por repetição, com fórmulas análogas a (11.46). A maior parte dos conjuntos tridimensionais que aparecem na prática são de um dos três tipos mencionados, ou podem decompor-se num número finito de partes cada uma das quais é de um desses tipos.

Existem muitas fórmulas de integração repetida para integrais n -múltiplos quando $n > 3$. Por exemplo, se Q é um intervalo k -dimensional e R um intervalo m -dimensional, então um integral $(m+k)$ -múltiplo sobre $Q \times R$ reduz-se à integração repetida de um m -múltiplo integral com um k -múltiplo integral,

$$\int_{Q \times R} \cdots \int f = \int_Q \cdots \int \left[\int_R \cdots \int f dx_1 \cdots dx_r \right] dx_{m+1} \cdots dx_{m+k},$$

com tanto que todos os integrais múltiplos intervenientes existam. Veremos ainda neste capítulo um exemplo do que se acaba de referir, ao calcularmos o volume de uma esfera no espaço n -dimensional.

11.32. Mudança de variáveis num integral n -múltiplo

A fórmula para a mudança de variáveis num integral duplo admite uma extensão directa para integrais n -múltiplos. Introduzamos novas variáveis u_1, \dots, u_n relacionadas com x_1, \dots, x_n por n equações da forma

$$x_1 = X_1(u_1, \dots, u_n), \quad \dots, \quad x_n = X_n(u_1, \dots, u_n).$$

Façamos $x = (x_1, \dots, x_m)$, $u = (u_1, \dots, u_n)$, e $X = (X_1, \dots, X_n)$. Então estas equações definem uma aplicação vectorial

$$X: T \rightarrow S$$

de um conjunto T do espaço n -dimensional noutro conjunto S do mesmo espaço. Suponhamos que a aplicação X é biunívoca e continuamente diferenciável em T . A fórmula de transformação por integrais n -múltiplos toma a forma

$$(11.47) \quad \int_S f(x) dx = \int_T f[X(u)] |\det DX(u)| du,$$

com $DX(u) = [D_j X_k(u)]$ a matriz jacobiana do campo vectorial X . Em função das componentes temos

$$DX(u) = \begin{bmatrix} D_1 X_1(u) & D_2 X_1(u) & \cdots & D_n X_1(u) \\ \vdots & \vdots & & \vdots \\ D_1 X_n(u) & D_2 X_n(u) & \cdots & D_n X_n(u) \end{bmatrix}.$$

Como no caso bidimensional, a fórmula de transformação é válida se X é biunívoca em T e se o jacobiano $J(u) = \det DX(u)$ nunca se anula em T . É ainda válida se a aplicação deixa apenas de ser biunívoca num subconjunto de T tendo medida n -dimensional nula, ou se o jacobiano se anula num tal subconjunto.

Para o caso tridimensional escrevemos (x, y, z) em vez de (x_1, x_2, x_3) , (u, v, w) em vez de (u_1, u_2, u_3) e (X, Y, Z) em vez de (X_1, X_2, X_3) . A fórmula de transformação para integrais triplos toma a forma

$$(11.48) \quad \iiint_S f(x, y, z) dx dy dz = \iiint_T f[X(u, v, w), Y(u, v, w), Z(u, v, w)] |J(u, v, w)| du dv dw,$$

onde $J(u, v, w)$ é o determinante jacobiano,

$$J(u, v, w) = \begin{vmatrix} \frac{\partial X}{\partial u} & \frac{\partial Y}{\partial u} & \frac{\partial Z}{\partial u} \\ \frac{\partial X}{\partial v} & \frac{\partial Y}{\partial v} & \frac{\partial Z}{\partial v} \\ \frac{\partial X}{\partial w} & \frac{\partial Y}{\partial w} & \frac{\partial Z}{\partial w} \end{vmatrix}.$$

No espaço tridimensional o determinante jacobiano pode considerar-se como um factor de ampliação de volumes. Com efeito, se introduzirmos a função vectorial r definida pela equação

$$\mathbf{r}(u, v, w) = X(u, v, w)\mathbf{i} + Y(u, v, w)\mathbf{j} + Z(u, v, w)\mathbf{k},$$

e os vectores

$$\mathbf{V}_1 = \frac{\partial \mathbf{r}}{\partial u} = \frac{\partial X}{\partial u} \mathbf{i} + \frac{\partial Y}{\partial u} \mathbf{j} + \frac{\partial Z}{\partial u} \mathbf{k},$$

$$\mathbf{V}_2 = \frac{\partial \mathbf{r}}{\partial v} = \frac{\partial X}{\partial v} \mathbf{i} + \frac{\partial Y}{\partial v} \mathbf{j} + \frac{\partial Z}{\partial v} \mathbf{k},$$

$$\mathbf{V}_3 = \frac{\partial \mathbf{r}}{\partial w} = \frac{\partial X}{\partial w} \mathbf{i} + \frac{\partial Y}{\partial w} \mathbf{j} + \frac{\partial Z}{\partial w} \mathbf{k},$$

um raciocínio semelhante ao dado na secção 11.26 sugere que um paralelepípedo rectângulo de arestas Δu , Δv , Δw no espaço $OUVW$ se transforma num sólido que é um quase "paralelepípedo" curvo do espaço $OXYZ$ definido pelos três vectores $\mathbf{V}_1 \Delta u$, $\mathbf{V}_2 \Delta v$ e $\mathbf{V}_3 \Delta w$. (Ver figura 11.30.) As fronteiras deste sólido são superfícies que se obtêm quando se faz respectivamente $u = \text{constante}$, $v = \text{constante}$ e $w = \text{constante}$. O volume de um paralelepípedo é igual ao valor absoluto do produto triplo escalar dos três vectores que o definem, pelo que o volume do paralelepípedo curvo será aproximadamente igual a

$$|(\mathbf{V}_1 \Delta u) \cdot (\mathbf{V}_2 \Delta v) \times (\mathbf{V}_3 \Delta w)| = |\mathbf{V}_1 \cdot \mathbf{V}_2 \times \mathbf{V}_3| \Delta u \Delta v \Delta w = |J(u, v, w)| \Delta u \Delta v \Delta w.$$

11.33. Exemplos resolvidos

Nos dois exemplos que apresentamos a seguir estudam-se dois casos particulares importantes de (11.48).

EXEMPLO 1. Coordenadas cilíndricas. Escrevemos aqui r , θ , z em vez de u , v , w e definimos a aplicação pelas equações

$$(11.49) \quad x = r \cos \theta, \quad y = r \sin \theta, \quad z = z.$$

Por outras palavras, substituímos x e y pelas suas coordenadas polares no plano XOY e deixamos z inalterado. De novo, para obtermos uma aplicação biunívoca devemos fazer $r > 0$ e restringir θ a valores de um dado intervalo da forma $\theta_1 \leq \theta \leq \theta_0 + 2\pi$. A figura 11.30 mostra o que acontece a um paralelepípedo rectângulo no espaço $r\theta z$.

O determinante jacobiano da aplicação em (11.49) é

$$J(r, \theta, z) = \begin{vmatrix} \cos \theta & \sin \theta & 0 \\ -r \sin \theta & r \cos \theta & 0 \\ 0 & 0 & 1 \end{vmatrix} = r(\cos^2 \theta + \sin^2 \theta) = r,$$

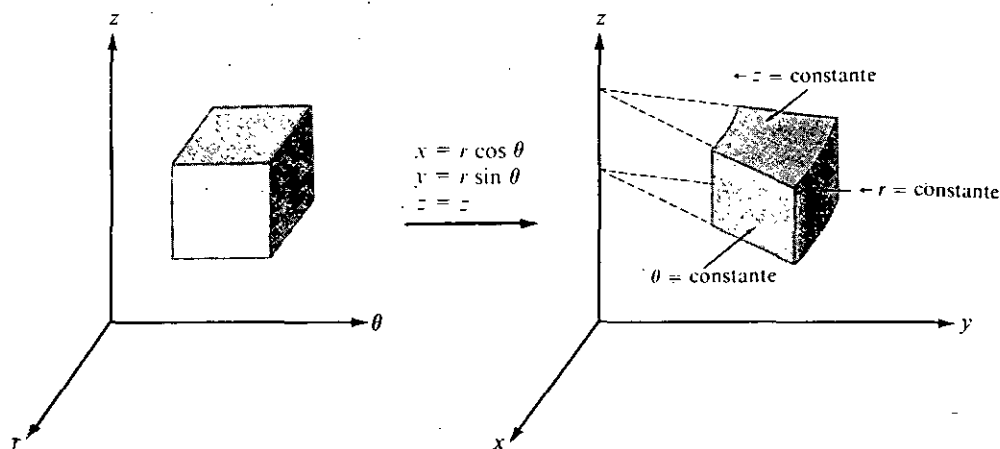


FIG. 11.30. Transformação por coordenadas cilíndricas.

e por conseguinte a fórmula de transformação em (11.48) vem

$$\iiint_S f(x, y, z) dx dy dz = \iiint_T f(r \cos \theta, r \sin \theta, z) r dr d\theta dz.$$

O determinante jacobiano anula-se quando $r = 0$, mas tal não afecta a validade da fórmula de transformação devido ao facto de o conjunto dos pontos com $r = 0$ ter medida nula.

EXEMPLO 2. Coordenadas esféricas. Neste caso usam-se os símbolos ρ , θ , φ em vez de u , v , w e a aplicação define-se pelas equações

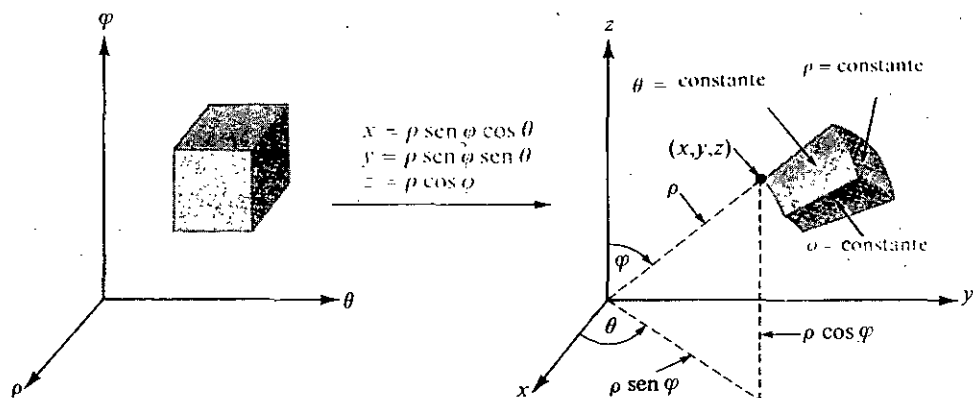


FIG. 11.31 Transformação por coordenadas esféricas

$$x = \rho \cos \theta \operatorname{sen} \varphi, \quad y = \rho \operatorname{sen} \theta \operatorname{sen} \varphi, \quad z = \rho \cos \varphi.$$

O significado geométrico de ρ , θ e φ está indicado na figura 11.31. Para obter uma aplicação biunívoca fazemos $\rho > 0$, $0 \leq \theta \leq 2\pi$ e $0 < \varphi < \pi$. As superfícies $\rho = \text{constante}$ são esferas centradas na origem, as superfícies $\theta = \text{constante}$ são planos passando pelo eixo OZ , e as superfícies $\varphi = \text{constante}$ são cones circulares com eixo coincidente com OZ . Portanto o paralelepípedo-rectângulo no espaço $0\rho\theta\varphi$ transforma-se num sólido da forma indicada na figura 11.31.

O determinante jacobiano é

$$J(\rho, \theta, \varphi) = \begin{vmatrix} \cos \theta \operatorname{sen} \varphi & \operatorname{sen} \theta \operatorname{sen} \varphi & \cos \varphi \\ -\rho \operatorname{sen} \theta \operatorname{sen} \varphi & \rho \cos \theta \operatorname{sen} \varphi & 0 \\ \rho \cos \theta \cos \varphi & \rho \operatorname{sen} \theta \cos \varphi & -\rho \operatorname{sen} \varphi \end{vmatrix} = -\rho^2 \operatorname{sen} \varphi.$$

Visto que $\operatorname{sen} \varphi \geq 0$ se $0 \leq \varphi < \pi$, temos $|J(\rho, \theta, \varphi)| = \rho^2 \operatorname{sen} \varphi$ e a fórmula de mudança de variáveis para integrais triplos vem

$$\iiint_S f(x, y, z) \, dx \, dy \, dz = \iiint_T F(\rho, \theta, \varphi) \rho^2 \operatorname{sen} \varphi \, d\rho \, d\theta \, d\varphi,$$

na qual se escreveu $F(\rho, \theta, \varphi)$ em vez de $f(\rho \cos \theta \operatorname{sen} \varphi, \rho \operatorname{sen} \theta \operatorname{sen} \varphi, \rho \cos \varphi)$. Embora o determinante jacobiano se anule quando $\varphi = 0$, a fórmula de mudança de variáveis é ainda válida porque o conjunto de pontos com $\varphi = 0$ tem medida nula.

O conceito de volume pode generalizar-se a certas classes de conjuntos (chamados conjuntos mensuráveis) num espaço n dimensional de tal maneira que se S é mensurável então o seu volume é igual ao integral da função constante 1 sobre S . Quer dizer, se $v(S)$ representa o volume de S , temos

$$v(S) = \int_S \cdots \int_S dx_1 \cdots dx_n.$$

Não tentaremos descrever a classe dos conjuntos para os quais esta fórmula é válida. Em vez disso, exporemos o modo como o integral pode calcular-se em alguns casos particulares.

EXEMPLO 3. *Volume de um intervalo n dimensional.* Se S é um intervalo n -dimensional, digamos $S = [a_1, b_1] \times \cdots \times [a_n, b_n]$, o integral múltiplo para o cálculo de $v(S)$ é o produto de n integrais unidimensionais,

$$v(S) = \int_{a_1}^{b_1} dx_1 \cdots \int_{a_n}^{b_n} dx_n = (b_1 - a_1) \cdots (b_n - a_n).$$

Isto coincide com a fórmula dada anteriormente para o volume de um intervalo n -dimensional.

EXEMPLO 4. O volume de uma esfera n -dimensional. Seja $S_n(a)$ a esfera n -dimensional (ou n -bola) de raio a definida por

$$S_n(a) = \{(x_1, \dots, x_n) \mid x_1^2 + \dots + x_n^2 \leq a^2\},$$

e seja

$$V_n(a) = \int_{S_n(a)} dx_1 \cdots dx_n,$$

o volume de $S_n(a)$. Vamos provar que

$$(11.50) \quad V_n(a) = \frac{\pi^{n/2}}{\Gamma(\frac{1}{2}n + 1)} a^n,$$

onde Γ é a função gama. Para $n = 1$ a fórmula dá $V_1(a) = 2a$, a amplitude do intervalo $[-a, a]$. Para $n = 2$ dá $V_2(a) = \pi a^2$, a área de um círculo de raio a . Vamos demonstrar (11.50) para $n \geq 3$.

Em primeiro lugar provamos que para cada $a > 0$ temos

$$(11.51) \quad V_n(a) = a^n V_n(1).$$

Por outras palavras, o volume de uma esfera de raio a é a^n vezes o volume de uma esfera de raio 1. Para demonstrarmos isto usamos uma mudança de variável linear $x = au$ para aplicar $S_n(1)$ em $S_n(a)$. A aplicação tem determinante jacobiano a^n . Logo

$$V_n(a) = \int_{S_n(a)} dx_1 \cdots dx_n = \int_{S_n(1)} a^n du_1 \cdots du_n = a^n V_n(1).$$

o que demonstra (11.51). Portanto, para provar (11.50) basta provar que

$$(11.52) \quad V_n(1) = \frac{\pi^{n/2}}{\Gamma(\frac{1}{2}n + 1)}.$$

Observe-se em primeiro lugar que $x_1^2 + \dots + x_n^2 \leq 1$ se e só se

$$x_1^2 + \dots + x_{n-2}^2 \leq 1 - x_{n-1}^2 - x_n^2 \quad \text{e} \quad x_{n-1}^2 + x_n^2 \leq 1.$$

Portanto podemos escrever o integral para $V_n(1)$ como uma integração repetida de um integral $(n-2)$ múltiplo e um integral duplo, como segue:

$$(11.53) \quad V_n(1) = \iint_{\substack{x_{n-1}^2 + x_n^2 \leq 1}} \left[\int \cdots \int_{\substack{x_1^2 + \dots + x_{n-2}^2 \leq 1 - x_{n-1}^2 - x_n^2}} dx_1 \cdots dx_{n-2} \right] dx_{n-1} dx_n.$$

O integral do parêntesis está estendido a uma esfera $S_{n-2}(R)$, onde $R = \sqrt{1 - x_{n-1}^2 - x_n^2}$, pelo que é igual a

$$V_{n-2}(R) = R^{n-2} V_{n-2}(1) = (1 - x_{n-1}^2 - x_n^2)^{n/2-1} V_{n-2}(1).$$

Escrevamos agora x em vez de x_{n-1} e y em vez de x_n . Então (11.53) vem

$$V_n(1) = V_{n-2}(1) \iint_{x^2+y^2 \leq 1} (1 - x^2 - y^2)^{n/2-1} dx dy.$$

Calculamos o integral duplo fazendo a mudança para coordenadas polares e obtemos

$$V_n(1) = V_{n-2}(1) \int_0^{2\pi} \int_0^1 (1 - r^2)^{n/2-1} r dr d\theta = V_{n-2}(1) \frac{2\pi}{n}.$$

Por outras palavras, os números $V_n(1)$ satisfazem à fórmula de recorrência

$$V_n(1) = \frac{2\pi}{n} V_{n-2}(1) \quad \text{se } n \geq 3.$$

Mas a sucessão de números $\{f(n)\}$ definida por

$$f(n) = \frac{\pi^{n/2}}{\Gamma(\frac{1}{2}n + 1)}$$

satisfaz à mesma fórmula de recorrência, porque $\Gamma(s+1) = s\Gamma(s)$. Também, $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ (ver Exercício 16, secção 11.28), pelo que $\Gamma(\frac{3}{2}) = \frac{1}{2}\sqrt{\pi}$ e $f(1) = V_1(1) = 2$. Também $f(2) = V_2(1) = \pi$, logo tem-se $f(n) = V_n(1)$ para todo $n \geq 1$, o que prova (11.52).

11.34. Exercícios

Calcular cada um dos integrais triplos dos Exercícios 1 a 5. Fazer, para cada um, um desenho da região de integração. Deve admitir-se a existência de todos os integrais encontrados.

1. $\iiint_S xy^2z^3 dx dy dz$, onde S é o sólido limitado pela superfície $z = xy$ e os planos $y = x$, $x = 1$, e $z = 0$.
2. $\iiint_S (1 + x + y + z)^3 dx dy dz$, onde S é o sólido limitado pelos três planos coordenados e o plano $x + y + z = 1$.
3. $\iiint_S xyz dx dy dz$, onde $S = \{(x, y, z) \mid x^2 + y^2 + z^2 \leq 1, x \geq 0, y \geq 0, z \geq 0\}$.
4. $\iiint_S \left(\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} \right) dx dy dz$, onde S é o sólido limitado por $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$.
5. $\iiint_S \sqrt{x^2 + y^2} dx dy dz$, onde S é o sólido formado pela folha superior do cone $z^2 = x^2 + y^2$ e o plano $z = 1$.

Nos Exercícios 6, 7 e 8 o integral triplo $\iiint_S f(x, y, z) dx dy dz$ de uma função positiva reduz-se à integração repetida que se indica. Para cada um deles descrever a região de integração S recorrendo a um desenho, mostrando a sua projecção no plano XOY . Expressar depois o integral triplo como um ou mais integrais repetidos nos quais a primeira integração se efectua em relação a y .

6. $\int_0^1 \left(\int_0^{1-x} \left[\int_0^{x+y} f(x, y, z) dz \right] dy \right) dx.$
7. $\int_{-1}^1 \left(\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \left[\int_{\sqrt{x^2-y^2}}^1 f(x, y, z) dz \right] dy \right) dx.$
8. $\int_0^1 \left(\int_0^1 \left[\int_0^{x^2+y^2} f(x, y, z) dz \right] dy \right) dx.$
9. Provar que

$$\int_0^x \left(\int_0^v \left[\int_0^u f(t) dt \right] du \right) dv = \frac{1}{2} \int_0^x (x-t)^2 f(t) dt.$$

Calcular os integrais dos Exercícios 10, 11, 12 por uma mudança para coordenadas cilíndricas. Admitir a existência de todos os integrais encontrados.

10. $\iiint (x^2 + y^2) dx dy dz$, onde S é o sólido limitado pela superfície $x^2 + y^2 = 2z$ e o plano $z = 2$.
11. $\iiint_S dx dy dz$, onde S é o sólido limitado pelos três planos coordenados, a superfície $z = x^2 + y^2$ e o plano $x + y = 1$.
12. $\iiint_S (y^2 + z^2) dx dy dz$, onde S é o cone circular recto de altura h e cuja base, de raio a , está situada no plano XOY e o eixo coincide com OZ .

Calcular os integrais nos Exercícios 13, 14 e 15 mediante uma mudança para coordenadas polares esféricas.

13. $\iiint_S dx dy dz$, onde S é a esfera de raio a e centro na origem.
14. $\iiint_S dx dy dz$, onde S é o sólido limitado por duas esferas concêntricas de raio a e b , ($0 < a < b$) e centro na origem.
15. $\iiint_S [(x-a)^2 + (y-b)^2 + (z-c)^2]^{-1/2} dx dy dz$, onde S é uma esfera de raio R e centro na origem, e (a, b, c) é um ponto fixo no exterior dessa esfera.
16. As coordenadas polares esféricas generalizadas podem definir-se pela seguinte aplicação

$$x = a \rho \cos^m \theta \sin^n \varphi, \quad y = b \rho \sin^m \theta \sin^n \varphi, \quad z = c \rho \cos^n \varphi,$$

onde a, b, c, m e n são constantes positivas. Mostrar que o jacobiano é igual a

$$-abcmn\rho^2 \cos^{m-1} \theta \sin^{m-1} \theta \cos^{n-1} \varphi \sin^{2n-1} \varphi.$$

Os integrais triplos podem ser utilizados para calcular volumes, massas, centros de massa, momentos de inércia e outros conceitos físicos relacionados com sólidos. Se S é um sólido, o seu volume V é dado pelo integral triplo

$$V = \iiint_S dx dy dz.$$

Se o sólido se supõe com densidade $f(x, y, z)$ em cada dos seus pontos (x, y, z) (massa por unidade de volumes), a sua massa M é dada por

$$M = \iiint_S f(x, y, z) \, dx \, dy \, dz,$$

e o respectivo centro de massa será o ponto $(\bar{x}, \bar{y}, \bar{z})$ em que

$$\bar{x} = \frac{1}{M} \iiint_S x f(x, y, z) \, dx \, dy \, dz,$$

e de modo análogo para \bar{y} e \bar{z} . O momento de inércia I_{xy} em relação ao plano XOY é definido por

$$I_{xy} = \iiint_S z^2 f(x, y, z) \, dx \, dy \, dz$$

e fórmulas semelhantes para I_{yz} e I_{zx} . O momento de inércia I_L em relação à recta L é definida por

$$I_L = \iiint_S \delta^2(x, y, z) f(x, y, z) \, dx \, dy \, dz,$$

em que $\delta^2(x, y, z)$ representa a distância do ponto genérico (x, y, z) de S à recta L .

17. Mostrar que os momentos de inércia em relação aos eixos coordenados são

$$I_x = I_{xy} + I_{xz}, \quad I_y = I_{yx} + I_{yz}, \quad I_z = I_{zx} + I_{zy}.$$

18. Determinar o volume do sólido limitado superiormente pela esfera $x^2 + y^2 + z^2 = 5$ e inferiormente pelo parabolóide $x^2 + y^2 = 4z$.
19. Determinar o volume do sólido limitado pelo plano XOY , o cilindro $x^2 + y^2 = 2x$ e o cone $z = \sqrt{x^2 + y^2}$.
20. Calcular a massa do sólido compreendido entre duas esferas concêntricas de raios a e b , (com $0 < a < b$), se a densidade em cada ponto for igual ao quadrado da distância deste ponto ao centro.
21. Um cone circular recto homogêneo tem altura h . Provar que a distância do seu centroide à base é $\frac{1}{4}h$.
22. Determinar o centro de massa de um cone circular recto de altura h , se a sua densidade em cada ponto é directamente proporcional à distância deste ponto à base.
23. Determinar o centro de massa de um cone circular recto de altura h , se a sua densidade em cada ponto é directamente proporcional à distância desse ponto ao eixo do cone.
24. Considere o sólido definido por duas semi-esferas concêntricas de raios a e b , com $0 < a < b$. Se a densidade for constante, determinar o centro de massa.
25. Determinar o centro de massa de um cubo de aresta h se a sua densidade em cada ponto for directamente proporcional ao quadrado da distância desse ponto a um vértice da base.
26. Um cone circular recto tem altura h , raio da base a , densidade constante, e massa M . Determinar o seu momento de inércia em relação a um eixo passando pelo vértice paralelo à base.
27. Determinar o momento de inércia de uma esfera de raio R e massa M relativamente a um diâmetro, supondo a sua densidade constante.

28. Determinar o momento de inércia de um cilindro de raio a e massa M se a sua densidade em cada ponto é proporcional à distância deste ponto ao eixo do cilindro.
29. A haste de um cogumelo é um cilindro circular recto de diâmetro 1 e altura 2 e a sua extremidade é uma semi-esfera de raio R . Se o cogumelo é considerado homogêneo com simetria axial e o seu centro de massa está situado no plano em que a haste se liga à extremidade, determinar R .
30. Um satélite artificial tem uma superfície exterior constituída por partes de dois cilindros circulares de iguais diâmetros D , cujos eixos se intersectam perpendicularmente. É necessário transportar o satélite para Cabo Kennedy numa caixa cúbica cuja aresta interior mede D . Provar que uma terça parte do volume da caixa fica vazia.
31. Seja $S_n(a)$ o seguinte conjunto no espaço n -dimensional, com $a > 0$:

$$S_n(a) = \{(x_1, \dots, x_n) \mid |x_1| + \dots + |x_n| \leq a\}.$$

Quando $n = 2$ o conjunto é um quadrado com vértices $(0, \pm a)$ e $(\pm a, 0)$.

Quando $n = 3$ é um octaedro com vértices $(0, 0, \pm a)$, $(0, \pm a, 0)$ e $(\pm a, 0, 0)$.

Seja $V_n(a)$ o volume de $S_n(a)$, dado por

$$V_n(a) = \int \cdots \int_{S_n(a)} dx_1 \cdots dx_n.$$

(a) Provar que $V_n(a) = a^n V_n(1)$.

(b) Para $n \geq 2$ exprimir o integral que dá $V_n(1)$ como uma integração repetida de um integral simples e outro integral $(n-1)$ -múltiplo e provar que

$$V_n(1) = V_{n-1}(1) \int_{-1}^1 (1 - |x|)^{n-1} dx = \frac{2}{n} V_{n-1}(1).$$

(c) Recorrendo às alíneas (a) e (b) provar que $V_n(a) = \frac{2^n a^n}{n!}$.

32. Seja $S_n(a)$ o seguinte conjunto no n -espaço, em que $a > 0$ e $n \geq 2$:

$$S_n(a) = \{(x_1, \dots, x_n) \mid |x_i| + |x_n| \leq a \text{ para cada } i = 1, \dots, n-1\}.$$

(a) Esboçar um desenho de $S_n(1)$ quando $n = 2$ e $n = 3$.

(b) Seja $V_n(a) = \int \cdots \int_{S_n(a)} dx_1 \cdots dx_n$, e provar que $V_n(a) = a^n V_n(1)$.

(c) Exprimir o integral que dá $V_n(1)$ como uma integração repetida de um integral simples e um integral $(n-1)$ múltiplo e deduzir que $V_n(a) = 2^n a^n / n$.

33. (a) Em relação com o Exemplo 4, pg. 462, exprimir o integral que dá $V_n(1)$, o volume da esfera n -dimensional de raio 1, como uma integração repetida de um integral $(n-1)$ múltiplo e um integral simples e com base nisso provar que

$$V_n(1) = 2V_{n-1}(1) \int_0^1 (1 - x^2)^{(n-1)/2} dx.$$

(b) Utilizar a alínea (a) e a equação (11.52) para provar que

$$\int_0^{\pi/2} \cos^n t \, dt = \frac{\sqrt{\pi}}{2} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2} + 1\right)}.$$

12

INTEGRAIS DE SUPERFÍCIE

12.1. Representação paramétrica de uma superfície

Este capítulo é dedicado ao estudo dos integrais de superfície e respectivas aplicações. Pode considerar-se o integral de superfície como o equivalente bidimensional do integral de linha, sendo a região de integração uma superfície em vez de uma curva. Antes que possamos estudar de maneira compreensível o que são os integrais de superfície, temos que estar de acordo sobre o que deve entender-se por uma superfície.

Grosseiramente falando, uma superfície é o lugar geométrico definido por um ponto que se move no espaço com dois graus de liberdade. No estudo de geometria analítica, efectuado no Volume I, analisámos dois métodos para definir tais lugares geométricos por meio de fórmulas matemáticas. Um é a *representação implícita*, na qual descrevemos uma superfície como um conjunto de pontos (x, y, z) verificando uma equação da forma $F(x, y, z) = 0$. Por vezes tal equação pode resolver-se em relação a uma das variáveis exprimindo-a em função das outras duas, seja por exemplo z em função de x e y . Quando tal é possível obtemos uma *representação explícita*, dada por uma ou mais equações da forma $z = f(x, y)$. Por exemplo, uma esfera de raio 1 e centro na origem tem a representação implícita $x^2 + y^2 + z^2 - 1 = 0$. Resolvendo em ordem a z obtemos duas soluções, $z = \sqrt{1 - x^2 - y^2}$ e $z = -\sqrt{1 - x^2 - y^2}$. A primeira dá-nos uma representação explícita da semi-esfera superior e a segunda uma representação do mesmo tipo para a semi-esfera inferior.

Existe um terceiro método de representação de superfícies, o qual é de maior utilidade no estudo dos integrais de superfície; trata-se da representação *vectorial* ou *paramétrica*, na qual três equações exprimem x , y , e z em função de dois parâmetros u e v .

$$(12.1) \quad x = X(u, v), \quad y = Y(u, v), \quad z = Z(u, v).$$

Aqui ao ponto (u, v) é permitido variar sobre um conjunto conexo bidimensional T do plano OUV , e os correspondentes pontos (x, y, z) constituem uma porção de superfície no espaço $OXYZ$. Este método de descrição de uma superfície é análogo à representação de uma curva, no espaço, por intermédio de três equações paramétricas a um parâmetro. A presença de dois parâmetros em (12.1) torna possível transmitir dois graus de liberdade ao ponto (x, y, z) , como se sugere na figura 12.1. Outra maneira de expressar a mesma ideia consiste em dizer que a superfície é a imagem de uma região plana T por intermédio da aplicação definida por (12.1).

Se introduzimos o raio vector \mathbf{r} da origem ao ponto genérico (x, y, z) da superfície, podemos combinar as três equações paramétricas (12.1) na equação vectorial da forma

$$(12.2) \quad \mathbf{r}(u, v) = X(u, v)\mathbf{i} + Y(u, v)\mathbf{j} + Z(u, v)\mathbf{k}, \quad \text{com } (u, v) \in T.$$

Esta é a chamada *equação vectorial* da superfície.

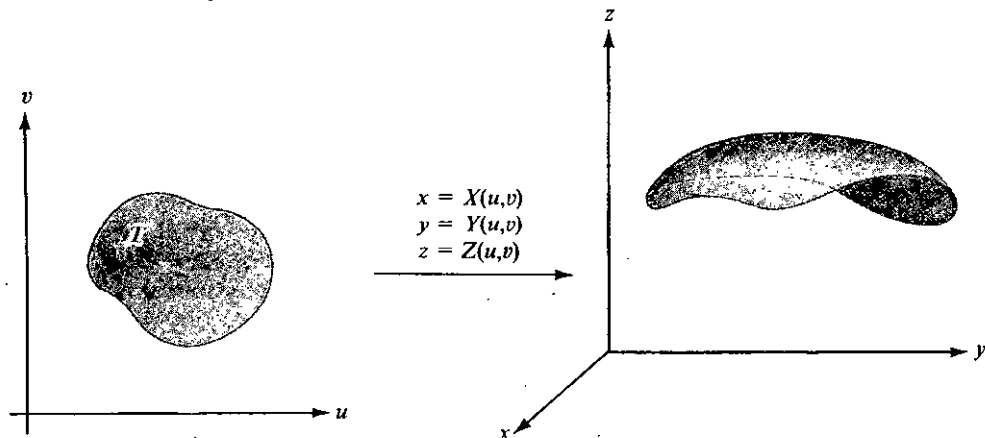


FIG. 12.1. Representação paramétrica de uma superfície.

Existem, naturalmente, muitas representações paramétricas para a mesma superfície. Uma delas pode sempre obter-se a partir da forma explícita $z = f(x, y)$, fazendo $X(u, v) = u$, $Y(u, v) = v$, $Z(u, v) = f(u, v)$. Por outro lado, se for possível resolver as duas primeiras equações (12.1) em relação a u e v em função de x e y e substituirmos na terceira obtemos a representação explícita $z = f(x, y)$.

EXEMPLO I. *Representação paramétrica da esfera.* As três equações

$$(12.3) \quad x = a \cos u \cos v, \quad y = a \sin u \cos v, \quad z = a \sin v$$

representam uma esfera de raio a e centro na origem. Quadrando e somando membro a membro as três equações (12.3) obtemos $x^2 + y^2 + z^2 = a^2$, concluindo assim que todo o ponto (x, y, z) que verifica (12.3) está situado sobre a esfera. Os parâmetros

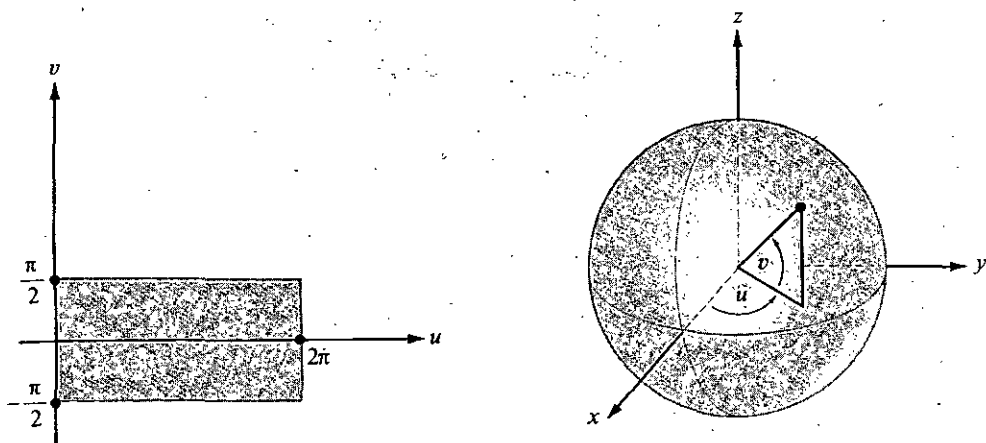


FIG. 12.2. Representação paramétrica de uma esfera.

u e v neste exemplo podem interpretar-se geometricamente como sendo os ângulos representados na figura 12.2. Se fizermos variar o ponto (u, v) no rectângulo $T = [0, 2\pi] \times [-\frac{\pi}{2}, \frac{\pi}{2}]$, os pontos definidos por (12.3) descrevem toda a esfera. O hemisfério superior e a imagem do rectângulo $[0, 2\pi] \times [0, \frac{\pi}{2}]$ e o inferior a do rectângulo $[0, 2\pi] \times [-\frac{\pi}{2}, 0]$. A figura 12.3 dá-nos uma ideia de como o rectângulo $[0, 2\pi] \times [0, \frac{\pi}{2}]$ é aplicado no hemisfério superior. Imaginemos que o rectângulo era construído em material plástico flexível capaz de esticar ou encolher. A figura 12.3 mostra o rectângulo a ser deformado numa semi-esfera. A base AB transforma-se aqui no equador, os lados apostos AD e BC são levados à coincidência, e o lado superior CD degenera num ponto (o polo norte).

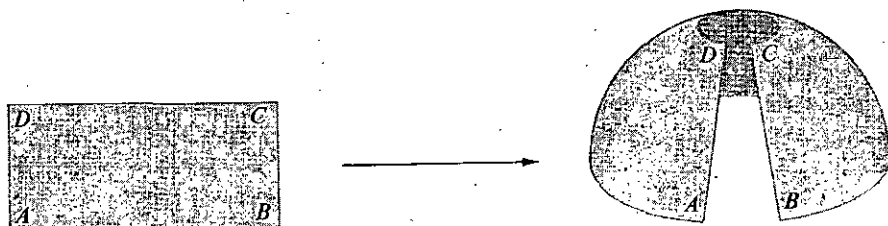


FIG. 12.3. Deformação de um rectângulo numa semi-esfera.

EXEMPLO 2. *Representação paramétrica de um cone.* A equação vectorial

$$\mathbf{r}(u, v) = v \sin \alpha \cos u \mathbf{i} + v \sin \alpha \sin u \mathbf{j} + v \cos \alpha \mathbf{k}$$

representa o cone circular recto da figura 12.4, na qual o ângulo α representa a semi-abertura do cone. Mais uma vez, podem os parâmetros u e v ser interpretados geometricamente; v é a distância do vértice ao ponto (x, y, z) do cone, e u o ângulo polar. Quando (u, v) varia no rectângulo $[0, 2\pi] \times [0, h]$, os pontos correspondentes (x, y, z) definem o cone de altura $h \cos \alpha$. Um rectângulo em matéria plástica pode, por

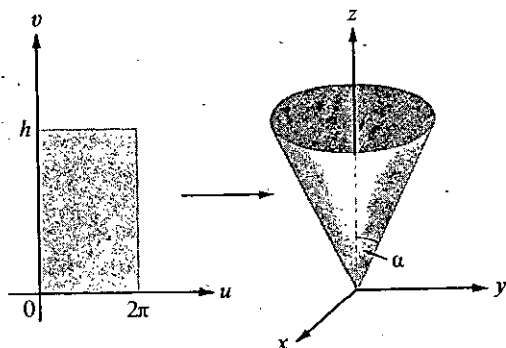


FIG. 12.4. Representação paramétrica do cone.

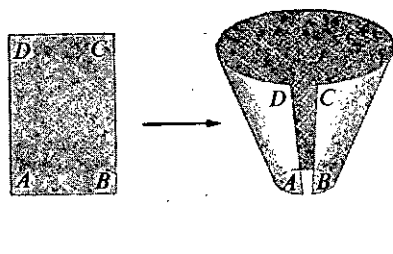


FIG. 12.5. Deformação de um rectângulo num cone.

deformação, transforma-se no cone obrigando os lados AD e BC a coincidirem, como se indica na figura 12.5, e fazendo com que o lado AB degenera num ponto (o vértice do cone). A superfície da figura 12.5 representa uma fase intermédia da deformação.

No estudo geral das superfícies, as funções X , Y e Z que aparecem nas equações paramétricas (12.1) ou na equação vectorial (12.2) supõem-se contínuas em T . A imagem de T atrás da aplicação \mathbf{r} chama-se uma *superfície paramétrica* e representa-se pelo símbolo $\mathbf{r}(T)$. Em muitos dos exemplos que apresentaremos, T será um rectângulo, um círculo, ou qualquer outro conjunto simplesmente conexo limitado por uma curva simples, fechada. Se a função \mathbf{r} é biunívoca em T , a imagem $\mathbf{r}(T)$ chama-se *superfície paramétrica simples* ou *superfície elementar*. Em tal hipótese, pontos distintos de T aplicam-se em pontos distintos da superfície. Em particular, toda a curva simples fechada em T aplica-se numa curva simples fechada pertencente à superfície.

Uma superfície paramétrica $\mathbf{r}(T)$ pode degenerar num ponto, ou numa curva. Por exemplo, se as três funções X , Y e Z são constantes, a imagem $\mathbf{r}(T)$ é um único ponto. Se X , Y e Z são independentes de v , a imagem $\mathbf{r}(T)$ é uma curva. Outro exemplo de superfície degenerada ocorre quando $X(u, v) = u + v$, $Y(u, v) = (u + v)^2$, e $Z(u, v) = (u + v)^3$, com $T = [0, 1] \times [0, 1]$. Se escrevermos $t = u + v$, vemos que a superfície degenera na curva de equações paramétricas $x = t$, $y = t^2$, e $z = t^3$, com $0 \leq t \leq 2$. Estes casos podem evitar-se impondo certas restrições à função \mathbf{r} que define a aplicação, como se explica na Secção seguinte.

12.2. O produto vectorial fundamental

Consideremos uma superfície definida pela equação vectorial

$$\mathbf{r}(u, v) = X(u, v)\mathbf{i} + Y(u, v)\mathbf{j} + Z(u, v)\mathbf{k}, \quad \text{com} \quad (u, v) \in T.$$

Se X, Y e Z são diferenciáveis em T , consideremos os dois vectores

$$\frac{\partial \mathbf{r}}{\partial u} = \frac{\partial X}{\partial u} \mathbf{i} + \frac{\partial Y}{\partial u} \mathbf{j} + \frac{\partial Z}{\partial u} \mathbf{k}$$

$$\frac{\partial \mathbf{r}}{\partial v} = \frac{\partial X}{\partial v} \mathbf{i} + \frac{\partial Y}{\partial v} \mathbf{j} + \frac{\partial Z}{\partial v} \mathbf{k}.$$

O produto vectorial destes dois vectores, $\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v}$, designar-se-á por *produto vectorial fundamental* da representação \mathbf{r} . As suas componentes podem exprimir-se por intermédio de determinantes jacobianos. Com efeito, temos

$$\begin{aligned} (12.4) \quad \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial X}{\partial u} & \frac{\partial Y}{\partial u} & \frac{\partial Z}{\partial u} \\ \frac{\partial X}{\partial v} & \frac{\partial Y}{\partial v} & \frac{\partial Z}{\partial v} \end{vmatrix} = \begin{vmatrix} \frac{\partial Y}{\partial u} & \frac{\partial Z}{\partial u} \\ \frac{\partial Y}{\partial v} & \frac{\partial Z}{\partial v} \end{vmatrix} \mathbf{i} + \begin{vmatrix} \frac{\partial Z}{\partial u} & \frac{\partial X}{\partial u} \\ \frac{\partial Z}{\partial v} & \frac{\partial X}{\partial v} \end{vmatrix} \mathbf{j} + \begin{vmatrix} \frac{\partial X}{\partial u} & \frac{\partial Y}{\partial u} \\ \frac{\partial X}{\partial v} & \frac{\partial Y}{\partial v} \end{vmatrix} \mathbf{k} \\ &= \frac{\partial(Y, Z)}{\partial(u, v)} \mathbf{i} + \frac{\partial(Z, X)}{\partial(u, v)} \mathbf{j} + \frac{\partial(X, Y)}{\partial(u, v)} \mathbf{k}. \end{aligned}$$

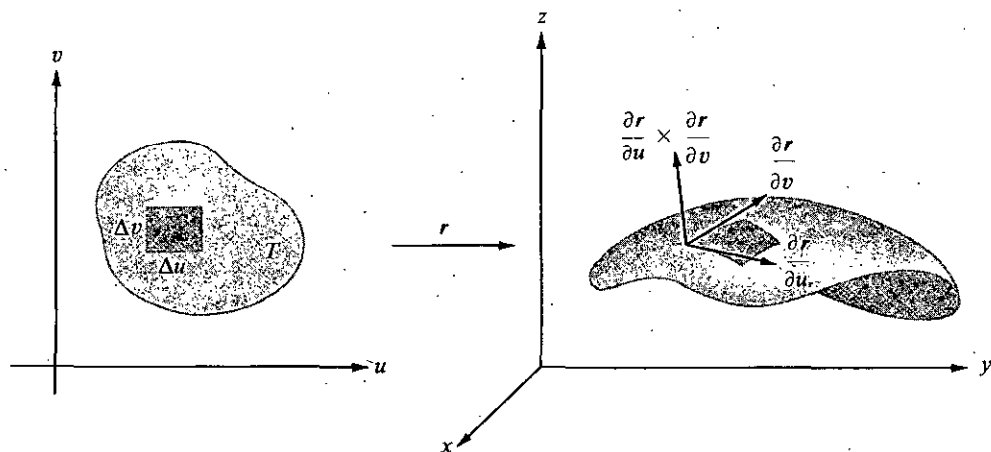


FIG. 12.6. Interpretação geométrica dos vectores $\frac{\partial \mathbf{r}}{\partial u}$, $\frac{\partial \mathbf{r}}{\partial v}$ e $\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v}$

Se (u, v) é um ponto de T no qual $\partial \mathbf{r}/\partial u$ e $\partial \mathbf{r}/\partial v$ são contínuas e o produto vectorial fundamental é não nulo; então o ponto imagem $\mathbf{r}(u, v)$ chama-se *ponto regular* de \mathbf{r} . Pontos nos quais $\partial \mathbf{r}/\partial u$ e $\partial \mathbf{r}/\partial v$ não são contínuos, ou em que $\partial \mathbf{r}/\partial u \times \partial \mathbf{r}/\partial v = \mathbf{0}$, dizem-se *pontos singulares* de \mathbf{r} . Uma superfície $\mathbf{r}(T)$ diz-se *regular* se todos os seus pontos o forem. Toda a superfície admite mais do que uma representação paramétrica. Alguns dos exemplos que analizaremos a seguir mostram que um ponto de uma superfície pode ser regular para uma determinada representação, mas singular noutra. O significado geométrico de pontos regulares e singulares pode explicar-se do modo seguinte:

Consideremos em T um segmento de recta horizontal. A sua imagem por \mathbf{r} é uma curva (chamada a curva u) pertencente à superfície $\mathbf{r}(T)$. Para um v fixo, consideremos o parâmetro u como representando o tempo. O vector $\partial \mathbf{r}/\partial u$ é o vector velocidade ao longo desta curva. Quando u varia de Δu , um ponto inicialmente situado em $\mathbf{r}(u, v)$ move-se ao longo de uma curva u de uma distância aproximadamente igual a $\|\partial \mathbf{r}/\partial u\| \Delta u$, uma vez que $\|\partial \mathbf{r}/\partial u\|$ representa a velocidade do ponto ao percorrer a curva $-u$. Analogamente, para u fixo, um ponto da curva v desloca-se, durante o intervalo de tempo Δv , de uma distância aproximadamente igual a $\|\partial \mathbf{r}/\partial v\| \Delta v$. Um rectângulo de T com área $\Delta v \Delta u$ transforma-se numa porção de $\mathbf{r}(T)$ que aproximaremos por um paralelogramo determinado pelos vectores $(\partial \mathbf{r}/\partial u) \Delta u$ e $(\partial \mathbf{r}/\partial v) \Delta v$. (Ver figura 12.6.) A área do paralelogramo determinado por $(\partial \mathbf{r}/\partial u) \Delta u$ e $(\partial \mathbf{r}/\partial v) \Delta v$ é o módulo do respectivo produto vectorial,

$$\left\| \frac{\partial \mathbf{r}}{\partial u} \Delta u \times \frac{\partial \mathbf{r}}{\partial v} \Delta v \right\| = \left\| \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right\| \Delta u \Delta v.$$

Deste modo a grandeza do produto vectorial fundamental pode considerar-se como um factor de ampliação de áreas. Nos pontos em que este produto vectorial é nulo o paralelogramo degenera numa curva ou num ponto. Em todo o ponto regular os vectores $\partial \mathbf{r}/\partial u$ e $\partial \mathbf{r}/\partial v$ determinam um plano cujo eixo é o vector $\partial \mathbf{r}/\partial u \times \partial \mathbf{r}/\partial v$. Na próxima Secção demonstraremos que $\partial \mathbf{r}/\partial u \times \partial \mathbf{r}/\partial v$ é normal a toda a curva regular da superfície; por este motivo o plano definido por $\partial \mathbf{r}/\partial u$ e $\partial \mathbf{r}/\partial v$ chama-se o *plano tangente* à superfície. A continuidade de $\partial \mathbf{r}/\partial u$ e $\partial \mathbf{r}/\partial v$ implica a continuidade de $\partial \mathbf{r}/\partial u \times \partial \mathbf{r}/\partial v$; esta, por sua vez, significa que o plano tangente varia continuamente para uma superfície regular. Vemos assim que a continuidade de $\partial \mathbf{r}/\partial u$ e $\partial \mathbf{r}/\partial v$ evita a ocorrência de arestas, ou pontos na superfície; o não anulamento de $\partial \mathbf{r}/\partial u \times \partial \mathbf{r}/\partial v$ assegura a não existência dos casos degenerados antes citados.

EXEMPLO 1. *Superfícies com uma representação explícita*, $z = f(x, y)$. Para uma superfície com uma representação explícita da forma $z = f(x, y)$, podemos usar x e y como parâmetros, o que nos conduz à equação vectorial

$$\mathbf{r}(x, y) = x\mathbf{i} + y\mathbf{j} + f(x, y)\mathbf{k}.$$

Esta representação conduz-nos sempre a uma superfície paramétrica simples. A região T diz-se a protecção da superfície sobre o plano XOY . (Na figura 12.7 está representado um exemplo.) Para calcular o produto vectorial fundamental observemos que

$$\frac{\partial \mathbf{r}}{\partial x} = \mathbf{i} + \frac{\partial f}{\partial x} \mathbf{k} \quad \text{e} \quad \frac{\partial \mathbf{r}}{\partial y} = \mathbf{j} + \frac{\partial f}{\partial y} \mathbf{k},$$

se f é diferenciável. Isto dá-nos

$$(12.5) \quad \frac{\partial \mathbf{r}}{\partial x} \times \frac{\partial \mathbf{r}}{\partial y} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 0 & \frac{\partial f}{\partial x} \\ 0 & 1 & \frac{\partial f}{\partial y} \end{vmatrix} = -\frac{\partial f}{\partial x} \mathbf{i} - \frac{\partial f}{\partial y} \mathbf{j} + \mathbf{k}.$$

Uma vez que a componente segundo OZ de $\partial \mathbf{r} / \partial x \times \partial \mathbf{r} / \partial y$ é 1, o produto vectorial fundamental nunca se anula. Portanto os únicos pontos singulares que podem aparecer nesta representação são pontos para os quais pelo menos uma das derivadas parciais $\partial f / \partial x$ ou $\partial f / \partial y$ deixa de ser contínua.

Um caso específico é o da equação $z = \sqrt{1 - x^2 - y^2}$, o qual representa uma semi-esfera de raio 1 e centro na origem, se $x^2 + y^2 \leq 1$. A equação vectorial

$$\mathbf{r}(x, y) = x\mathbf{i} + y\mathbf{j} + \sqrt{1 - x^2 - y^2} \mathbf{k}$$

aplica o círculo de raio 1, $T = \{(x, y) \mid x^2 + y^2 \leq 1\}$ sobre a semi-esfera, segundo uma aplicação biunívoca. As derivadas parciais $\partial \mathbf{r} / \partial x$ e $\partial \mathbf{r} / \partial y$ existem e são contínuas em todo o interior do círculo, mas não existem sobre a sua fronteira. Consequentemente cada ponto do equador é um ponto singular desta representação.

EXEMPLO 2. Consideremos a mesma semi-esfera do Exemplo 1, mas desta vez como a imagem do rectângulo $T = [0, 2\pi] \times [0, \frac{\pi}{2}]$ através da aplicação

$$\mathbf{r}(u, v) = a \cos u \cos v \mathbf{i} + a \sin u \cos v \mathbf{j} + a \sin v \mathbf{k}.$$

Os vectores $\partial \mathbf{r} / \partial u$ e $\partial \mathbf{r} / \partial v$ são dados por

$$\frac{\partial \mathbf{r}}{\partial u} = -a \sin u \cos v \mathbf{i} + a \cos u \cos v \mathbf{j},$$

$$\frac{\partial \mathbf{r}}{\partial v} = -a \cos u \sin v \mathbf{i} - a \sin u \sin v \mathbf{j} + a \cos v \mathbf{k}.$$

Um cálculo fácil mostra que o respectivo produto vectorial é igual a

$$\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} = a \cos v \mathbf{r}(u, v).$$

A imagem de T não é uma superfície paramétrica simples porque esta aplicação não é biunívoca em T . Com efeito, cada ponto do segmento de recta $v = \frac{1}{2}\pi$, $0 \leq u \leq 2\pi$, é aplicado no ponto $(0, 0, a)$ (o polo norte). Também, devido à periodicidade do seno e cosseno, r toma os mesmos valores nos pontos $(0, v)$ e $(2\pi, v)$, pelo que os lados direito e esquerdo de T são aplicados sobre a mesma curva, um arco de circunferência que une o polo norte ao ponto $(a, 0, 0)$ do equador. (Ver figura 12.3.) Os vectores $\partial r/\partial u$ e $\partial r/\partial v$ são contínuos em todo o T . Uma vez que $\|\partial r/\partial u \times \partial r/\partial v\| = a^2 \cos v$, os únicos pontos singulares desta representação aparecem quando $\cos v = 0$. O polo norte é o único de tais pontos.

12.3. O produto vectorial fundamental definindo uma normal à superfície

Consideremos uma superfície paramétrica regular $r(T)$, e seja C^* uma curva regular em T . Então a imagem $C = r(C^*)$ é uma curva regular situada na superfície. Provaremos que em cada ponto de C o vector $\partial r/\partial u \times \partial r/\partial v$ é normal a C , como indicado na figura 12.6.

Suponhamos que C^* é definida por uma função α definida num intervalo $[a, b]$, seja

$$\alpha(t) = U(t)i + V(t)j.$$

Então a imagem da curva C é representada por uma função composta

$$\rho(t) = r[\alpha(t)] = X[\alpha(t)]i + Y[\alpha(t)]j + Z[\alpha(t)]k.$$

Pretendemos provar que a derivada $\rho'(t)$ é perpendicular ao vector $\partial r/\partial u \times \partial r/\partial v$, com as derivadas parciais $\partial r/\partial u$ e $\partial r/\partial v$ calculadas em $(U(t), V(t))$. Para calcular $\rho'(t)$ derivamos cada componente de $\rho(t)$ (teorema 8.8) e obteremos

$$(12.6) \quad \rho'(t) = \nabla X \cdot \alpha'(t)i + \nabla Y \cdot \alpha'(t)j + \nabla Z \cdot \alpha'(t)k,$$

onde os vectores gradientes ∇X , ∇Y e ∇Z se calculam em $(U(t), V(t))$. A equação (12.6) pode escrever-se na forma

$$\rho'(t) = \frac{\partial r}{\partial u} U'(t) + \frac{\partial r}{\partial v} V'(t),$$

com as derivadas $\partial r/\partial u$ e $\partial r/\partial v$ calculadas em $(U(t), V(t))$. Uma vez que $\partial r/\partial u$ e $\partial r/\partial v$ são, cada um deles, perpendiculares ao produto vectorial $\partial r/\partial u \times \partial r/\partial v$, o mesmo se pode afirmar para $\rho'(t)$. Isto prova que $\partial r/\partial u \times \partial r/\partial v$ é normal a C , como tínhamos afirmado. Por este motivo, o vector $\partial r/\partial u \times \partial r/\partial v$ diz-se ser *normal* à superfície $r(T)$. Em cada ponto regular P de $r(T)$ o vector $\partial r/\partial u \times \partial r/\partial v$ é diferente de zero; o plano que passa por P e admite este vector como eixo chama-se o *plano tangente* à superfície em P .

12.4. Exercícios

Nos Exercícios 1 a 6, eliminar os parâmetros u e v a fim de se obter a equação cartesiana, provando que a equação vectorial dada representa a superfície que se indica. Calcular, também, o produto vectorial fundamental $\partial \mathbf{r} / \partial u \times \partial \mathbf{r} / \partial v$ em função de u e v .

1. Plano

$$\mathbf{r}(u, v) = (x_0 + a_1 u + b_1 v)\mathbf{i} + (y_0 + a_2 u + b_2 v)\mathbf{j} + (z_0 + a_3 u + b_3 v)\mathbf{k}.$$

2. Paraboloide elíptico:

$$\mathbf{r}(u, v) = au \cos v \mathbf{i} + bu \sin v \mathbf{j} + u^2 \mathbf{k}.$$

3. Elipsoide:

$$\mathbf{r}(u, v) = a \sin u \cos v \mathbf{i} + b \sin u \sin v \mathbf{j} + c \cos u \mathbf{k}.$$

4. Superfície de revolução:

$$\mathbf{r}(u, v) = u \cos v \mathbf{i} + u \sin v \mathbf{j} + f(u)\mathbf{k}.$$

5. Cilindro:

$$\mathbf{r}(u, v) = u\mathbf{i} + a \sin v \mathbf{j} + a \cos v \mathbf{k}.$$

6. Toro:

$$\mathbf{r}(u, v) = (a + b \cos u) \sin v \mathbf{i} + (a + b \cos u) \cos v \mathbf{j} + b \sin u \mathbf{k}, \text{ em que } 0 < b < a. \text{ Qual o significado geométrico de } a \text{ e } b?$$

Nos Exercícios 7 a 10 calcular a grandeza do produto vectorial $\partial \mathbf{r} / \partial u \times \partial \mathbf{r} / \partial v$ em função de u e v .

$$7. \mathbf{r}(u, v) = a \sin u \cosh v \mathbf{i} + b \cos u \cosh v \mathbf{j} + c \sinh v \mathbf{k}.$$

$$8. \mathbf{r}(u, v) = (u + v)\mathbf{i} + (u - v)\mathbf{j} + 4v^2 \mathbf{k}.$$

$$9. \mathbf{r}(u, v) = (u + v)\mathbf{i} + (u^2 + v^2)\mathbf{j} + (u^3 + v^3)\mathbf{k}.$$

$$10. \mathbf{r}(u, v) = u \cos v \mathbf{i} + u \sin v \mathbf{j} + \frac{1}{2} u^2 \sin 2v \mathbf{k}.$$

12.5. Área de uma superfície na representação paramétrica

Seja $S = \mathbf{r}(T)$ uma superfície paramétrica representada por uma função vectorial \mathbf{r} definida na região T do plano OUV . Na Secção 12.2 vimos que a grandeza do produto vectorial fundamental $\partial \mathbf{r} / \partial u \times \partial \mathbf{r} / \partial v$ pode ser interpretada como um factor de ampliação de áreas. (Ver figura 12.6.) Um rectângulo em T de área $\Delta u \Delta v$ transforma-se, por intermédio de \mathbf{r} num paralelogramo curvilíneo de S com área aproximadamente igual a

$$\left\| \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right\| \Delta u \Delta v.$$

Esta observação sugere a seguinte definição

DEFINIÇÃO DE ÁREA DE UMA SUPERFÍCIE NA REPRESENTAÇÃO PARAMÉTRICA. A área de S , representada por $\mathbf{r}(S)$, define-se pelo integral duplo

$$(12.7) \quad a(S) = \iint_T \left\| \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right\| du dv.$$

Por outras palavras, para determinar a área de S calculamos em primeiro lugar o produto vectorial fundamental $\partial \mathbf{r} / \partial u \times \partial \mathbf{r} / \partial v$ e integramos depois o seu comprimento sobre a região T . Quando $\partial \mathbf{r} / \partial u \times \partial \mathbf{r} / \partial v$ está expresso em função das suas componentes, por intermédio da equação (12.4), obtemos

$$(12.8) \quad a(S) = \iint_T \sqrt{\left(\frac{\partial(Y, Z)}{\partial(u, v)}\right)^2 + \left(\frac{\partial(Z, X)}{\partial(u, v)}\right)^2 + \left(\frac{\partial(X, Y)}{\partial(u, v)}\right)^2} du dv.$$

Escrito nesta forma, o integral para a área de uma superfície assemelha-se ao integral para o cálculo do comprimento de arco de uma curva.†

Se S é dada explicitamente por uma equação da forma $z = f(x, y)$, podemos usar x e y para parâmetros. O produto vectorial fundamental é dado pela equação (12.5), pelo que se tem

$$\left\| \frac{\partial \mathbf{r}}{\partial x} \times \frac{\partial \mathbf{r}}{\partial y} \right\| = \left\| -\frac{\partial f}{\partial x} \mathbf{i} - \frac{\partial f}{\partial y} \mathbf{j} + \mathbf{k} \right\| = \sqrt{1 + \left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}.$$

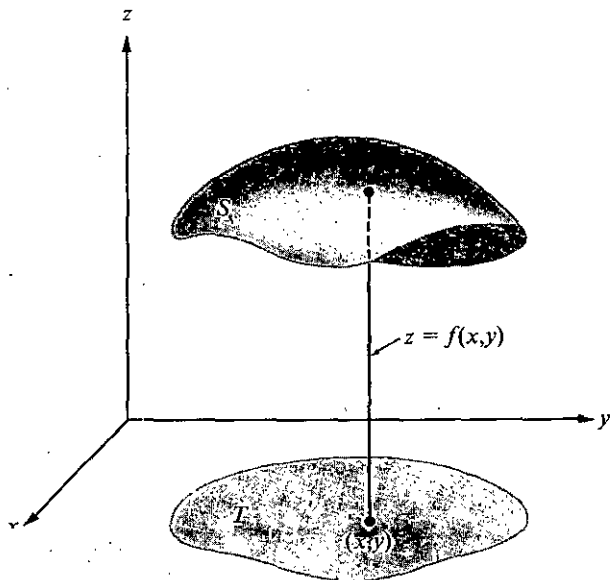


FIG. 12.7. Uma superfície S com uma representação explícita, $z = f(x, y)$. A região T é a projecção de S no plano XOY .

† Uma vez que o integral (12.7) contém r , a área de uma superfície dependerá da função usada para representar a superfície. Quando analisarmos os integrais de superfície provaremos (Secção 12.8) que sob certas condições gerais a área é independente da representação paramétrica. O resultado é análogo ao teorema 10.1, no qual analisámos a invariância dos integrais de linha perante uma mudança do parâmetro.

Neste caso o integral de superfície vem

$$(12.9) \quad a(S) = \iint_T \sqrt{1 + \left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2} dx dy,$$

onde a região T é agora a projecção de S sobre o plano XOY , como se mostra na figura 12.7.

Quando S está num plano paralelo ao plano XOY , a função f é constante, pelo que $\partial f/\partial x = \partial f/\partial y = 0$, e a equação (12.9) vem então

$$a(S) = \iint_T dx dy.$$

o que está de acordo com a fórmula usual para o cálculo de áreas de regiões planas.

A equação (12.9) pode escrever-se de outra maneira a qual nos fará compreender melhor o seu significado geométrico. Em cada ponto de S , represente γ o ângulo entre o vector normal a S , $N = \partial \mathbf{r}/\partial x \times \partial \mathbf{r}/\partial y$, e o vector unitário coordenado k . (Ver figura 12.8.) Uma vez que a componente segundo OZ de N é 1, temos

$$\cos \gamma = \frac{N \cdot k}{\|N\| \|k\|} = \frac{1}{\|N\|} = \frac{1}{\left\| \frac{\partial \mathbf{r}}{\partial x} \times \frac{\partial \mathbf{r}}{\partial y} \right\|},$$

e consequentemente $\left\| \frac{\partial \mathbf{r}}{\partial x} \times \frac{\partial \mathbf{r}}{\partial y} \right\| = 1/\cos \gamma$. Portanto a equação (12.9) vem

$$(12.10) \quad a(S) = \iint_T \frac{1}{\cos \gamma} dx dy.$$

Suponhamos agora que S está situada num plano não perpendicular ao plano XOY . Então γ é constante e a equação (12.10) estabelece que a área de $S = (\text{área de } T)/\cos \gamma$, ou que

$$(12.11) \quad a(T) = a(S) \cos \gamma.$$

A equação (12.11) designa-se algumas vezes por *princípio do cosseno para as áreas*. Diz-nos que se uma região S num plano é projectada sobre uma região T noutro plano, sendo γ o ângulo dos dois planos, a área de T é igual a $\cos \gamma$ vezes a área de S . Esta fórmula é, evidentemente, verdadeira quando S é o rectângulo indicado na figura 12.9, porque as distâncias numa direcção são encurtadas por um factor $\cos \gamma$, enquanto que as consideradas na direcção perpendicular permanecem inalteradas na mesma projecção. A equação (12.11) estende esta propriedade a qualquer região plana S .

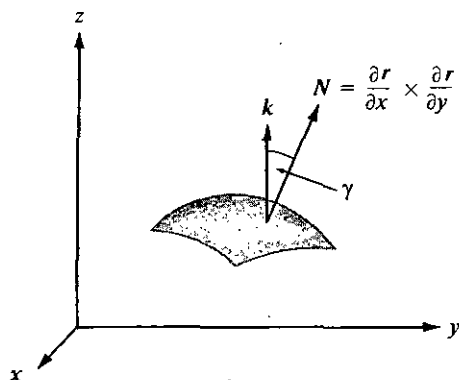


FIG. 12.8. A grandeza de $\frac{\partial \mathbf{r}}{\partial x} \times \frac{\partial \mathbf{r}}{\partial y}$ é $1/\cos \gamma$.

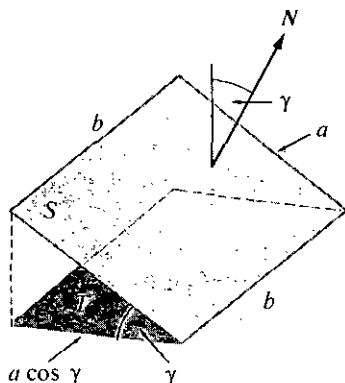


FIG. 12.9. O princípio do cosseno para as áreas para o caso de um rectângulo.

Suponhamos agora que S é dada por uma equação da forma $F(x, y, z) = 0$. Se S puder projectar-se da uma forma biunívoca sobre o plano XOY , a equação $F(x, y, z) = 0$ define z como função de x e y , isto é $z = f(x, y)$ e as derivadas parciais $\partial f/\partial x$ e $\partial f/\partial y$ estão relacionadas com as de F pelas igualdades

$$\frac{\partial f}{\partial x} = -\frac{\partial F/\partial x}{\partial F/\partial z} \quad \text{e} \quad \frac{\partial f}{\partial y} = -\frac{\partial F/\partial y}{\partial F/\partial z}$$

nos pontos em que $\partial F/\partial z \neq 0$. Substituindo estes cocientes em (12.9), encontramos

$$(12.12) \quad a(S) = \iint_T \frac{\sqrt{(\partial F/\partial x)^2 + (\partial F/\partial y)^2 + (\partial F/\partial z)^2}}{|\partial F/\partial z|} dx dy.$$

EXEMPLO 1. Área de uma semi-esfera. Consideremos uma semi-esfera S de raio a e centro na origem. Dispomos de imediato da representação implícita $x^2 + y^2 + z^2 = a^2$, $z \geq 0$; a representação explícita é $z = \sqrt{a^2 - x^2 - y^2}$; e a representação paramétrica é

$$(12.13) \quad \mathbf{r}(u, v) = a \cos u \cos v \mathbf{i} + a \sin u \cos v \mathbf{j} + a \sin v \mathbf{k}.$$

Para calcular a área de S a partir da representação implícita vamos utilizar (12.12) com

$$F(x, y, z) = x^2 + y^2 + z^2 - a^2.$$

As derivadas parciais de F são $\partial F/\partial x = 2x$, $\partial F/\partial y = 2y$, $\partial F/\partial z = 2z$. A semi-esfera S projecta-se, numa correspondência biunívoca, sobre o círculo $D = \{(x, y)/x^2 + y^2 \leq a^2\}$ no plano XOY . Não podemos aplicar (12.12) directamente porque a derivada parcial $\partial F/\partial z$ é zero sobre a fronteira de D . Contudo, a derivada $\partial F/\partial z$ é não nula em todo

o interior de D pelo que podemos considerar um círculo menor concêntrico com o primeiro $D(R)$ de raio R , com $R < a$. Se $S(R)$ representa a parte correspondente ao hemisfério superior, a equação (12.12) é agora aplicável e encontramos

$$\begin{aligned}\text{área de } S(R) &= \iint_{D(R)} \frac{\sqrt{(2x)^2 + (2y)^2 + (2z)^2}}{|2z|} dx dy \\ &= \iint_{D(R)} \frac{a}{z} dx dy = a \iint_{D(R)} \frac{1}{\sqrt{a^2 - x^2 - y^2}} dx dy.\end{aligned}$$

O último integral pode calcular-se facilmente usando coordenadas polares, obtendo-se:

$$\text{área de } S(R) = a \int_0^{2\pi} \left[\int_0^R \frac{1}{\sqrt{a^2 - r^2}} r dr \right] d\theta = 2\pi a(a - \sqrt{a^2 - R^2}).$$

Quando $R \rightarrow a$ a área de $S(R)$ tende para o limite $2\pi a^2$.

Podemos evitar a passagem ao limite no exemplo precedente recorrendo à representação paramétrica dada em (12.13). Os cálculos do Exemplo 2 da Secção mostram que

$$\left\| \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right\| = \|a \cos v \mathbf{r}(u, v)\| = a^2 |\cos v|.$$

Deste modo podemos aplicar a equação (12.7), tomando para a região T o rectângulo $[0, 2\pi] \times [0, \frac{1}{2}\pi]$. Encontramos

$$a(S) = a^2 \iint_T |\cos v| du dv = a^2 \int_0^{2\pi} \left[\int_0^{\pi/2} \cos v dv \right] du = 2\pi a^2.$$

EXEMPLO 2. Outro teorema de Pappus. Um dos teoremas de Pappus estabelece que a superfície de revolução, obtida por rotação de uma curva plana, de comprimento L , em torno de um eixo do seu plano, tem a área $2\pi Lh$, sendo h a distância ao eixo de rotação do centroide da curva. Usaremos a fórmula (12.7) para demonstrar este teorema.

Suponhamos uma curva C , inicialmente no plano XOY , que roda em torno do eixo OZ . Seja $z = f(x)$ a equação dessa curva no plano XOZ , com $a \leq x \leq b$, $a \geq 0$. A superfície de revolução S assim gerada pode definir-se pela equação vectorial

$$\mathbf{r}(u, v) = u \cos v \mathbf{i} + u \sin v \mathbf{j} + f(u) \mathbf{k},$$

onde $(u, v) \in [a, b] \times [0, 2\pi]$. Os parâmetros u e v podem interpretar-se como o raio vector e o ângulo polar de um sistema de coordenadas polares, como se indica na figura 12.10. Se $a \leq u \leq b$, todos os pontos (x, y, z) a uma dada distância u do eixo OZ têm a mesma cota, $f(u)$, pelo que estão todos sobre a superfície. O produto vectorial fundamental desta representação é

$$\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \cos v & \sin v & f'(u) \\ -u \sin v & u \cos v & 0 \end{vmatrix} = -uf'(u) \cos v \mathbf{i} - uf'(u) \sin v \mathbf{j} + u \mathbf{k},$$

e por isso

$$\left\| \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right\| = u \sqrt{1 + [f'(u)]^2}.$$

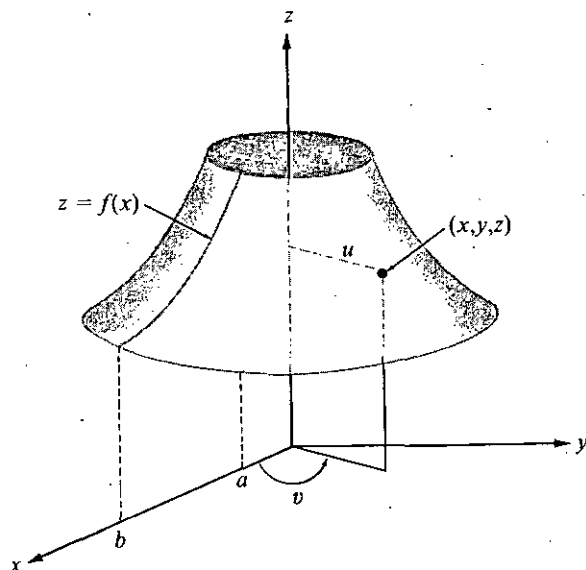


FIG. 12.10. Área de uma superfície de revolução determinada por aplicação do teorema de Pappus.

Deste modo a equação (12.7) escreve-se

$$a(S) = \int_0^{2\pi} \left[\int_a^b u \sqrt{1 + [f'(u)]^2} du \right] dv = 2\pi \int_a^b u \sqrt{1 + [f'(u)]^2} du.$$

O último integral pode exprimir-se na forma $\int_C x ds$, um integral de linha a respeito do comprimento de arco tomado ao longo de C . Como tal, é igual a $\bar{x}L$, sendo \bar{x} a abcissa do centroide de C e L o comprimento de C . (Ver Secção 10.8.) Consequentemente a área de S é $2\pi L\bar{x}$, o que demonstra o teorema de Pappus.

12.6. Exercícios

1. Seja S um paralelogramo não paralelo a qualquer dos planos coordenados. Sejam S_1 , S_2 e S_3 as áreas das projecções de S sobre os três planos coordenados. Provar que a área de S é $\sqrt{S_1^2 + S_2^2 + S_3^2}$.
2. Calcular a área da região que o plano $x + y + z = a$ determina no cilindro $x^2 + y^2 = a^2$.
3. Calcular a área da superfície da esfera $x^2 + y^2 + z^2 = a^2$, situada no interior do cilindro $x^2 + y^2 = ay$, com $a > 0$.
4. Calcular a área da porção da superfície $z^2 = 2xy$ situada acima do primeiro quadrante do plano XOY e limitado pelos planos $x = 2$ e $y = 1$.
5. Uma superfície S admite a representação paramétrica

$$\mathbf{r}(u, v) = u \cos v \mathbf{i} + u \sin v \mathbf{j} + u^2 \mathbf{k},$$

com $0 \leq u \leq 4$ e $0 \leq v \leq 2\pi$.

- (a) Provar que S é uma parte de uma superfície de revolução. Fazer um desenho e indicar o significado geométrico dos parâmetros u e v na superfície.
 - (b) Calcular o producto vectorial fundamental $\partial \mathbf{r} / \partial u \times \partial \mathbf{r} / \partial v$ em função de u e v .
 - (c) A área de S é $\pi(65\sqrt{65} - 1)/n$, com n inteiro. Calcular o valor de n .
6. Calcular a área da porção da superfície cónica $x^2 + y^2 = z^2$ situada acima de XOY e limitada pela esfera $x^2 + y^2 + z^2 = 2ax$.
 7. Calcular a área da porção da superfície cónica $x^2 + y^2 = z^2$ que está situada entre os planos $z = 0$ e $x + 2z = 3$.
 8. Calcular a área da porção do parabolóide $x^2 + z^2 = 2ay$ cortada pelo plano $y = a$.
 9. Calcular a área do toro de equação vectorial.

$$\mathbf{r}(u, v) = (a + b \cos u) \sin v \mathbf{i} + (a + b \cos u) \cos v \mathbf{j} + b \sin u \mathbf{k},$$

em que $0 < b < a$ e $0 \leq u \leq 2\pi$, $0 \leq v \leq 2\pi$. [Sugestão: Usar o teorema de Pappus.]

10. Uma esfera está inscrita num cilindro circular recto. A esfera é cortada por dois planos paralelos perpendiculares ao eixo do cilindro. Demonstrar que as partes da esfera e do cilindro compreendidas entre esses dois planos têm a mesma área.
11. Sejam T o círculo de raio e unidade no plano OUV ,

$$\mathbf{r}(u, v) = \frac{2u}{u^2 + v^2 + 1} \mathbf{i} + \frac{2v}{u^2 + v^2 + 1} \mathbf{j} + \frac{u^2 + v^2 - 1}{u^2 + v^2 + 1} \mathbf{k}.$$

- (a) Determinar a imagem, por \mathbf{r} , de cada uma dos seguintes conjuntos: o círculo unitário $u^2 + v^2 = 1$; o intervalo $-1 \leq u \leq 1$; a parte da recta $u = r$ situada em T .
- (b) A superfície $S = \mathbf{r}(T)$ é muito conhecida. Indicar o seu nome e desenhá-la.
- (c) Determinar a imagem por \mathbf{r} do plano OUV . Indicar por meio de um desenho qual o significado geométrico dos parâmetros u e v .

12.7. Integrais de superfície

Os integrais de superfície são, em muitos aspectos, análogos aos integrais de linha: a integração é estendida a uma superfície em vez de uma curva. Definimos os integrais de linha mediante uma representação paramétrica para a curva. Analogamente, definiremos integrais de superfície por intermédio de uma representação paramétrica

da superfície. Teremos então que provar que, sob certas condições gerais, o valor do integral é independente da representação considerada.

DEFINIÇÃO DE INTEGRAL DE SUPERFÍCIE. Seja $S = \mathbf{r}(t)$ uma superfície na representação paramétrica descrita por uma função diferenciável \mathbf{r} definida numa região T do plano u, v , e seja f um campo escalar definido e limitado em S . O integral de superfície de f sobre S representa-se pelo símbolo $\iint_{\mathbf{r}(T)} f dS$ ou por $\iint_S f(x, y, z) dS$ e define-se pela igualdade

$$(12.14) \quad \iint_{\mathbf{r}(T)} f dS = \iint_T f[\mathbf{r}(u, v)] \left\| \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right\| du dv$$

sempre que o integral duplo do segundo membro existe.

Os exemplos que se seguem ilustram algumas aplicações dos integrais de superfície.

EXEMPLO 1. Área de uma superfície. Quando $f = 1$, a equação (12.14) escreve-se

$$\iint_{\mathbf{r}(T)} dS = \iint_T \left\| \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right\| du dv.$$

O integral duplo do segundo membro é o que foi usado atrás na Secção (12.5) para definir área de uma superfície. Assim, a área de S é igual ao integral de superfície $\iint_{\mathbf{r}(T)} dS$. Por tal motivo, o símbolo dS chama-se por vezes “elemento de área da superfície” e o integral de superfície $\iint_{\mathbf{r}(T)} f dS$ diz-se ser um integral de f em relação ao elemento de área da superfície, estendido a toda a superfície $\mathbf{r}(T)$.

EXEMPLO 2. Centro de massa. Momento de inércia. Se num campo escalar, f é interpretado como sendo a densidade de massa de uma lâmina delgada com a forma de S , a massa total m da superfície obtém-se pela equação

$$m = \iint_S f(x, y, z) dS.$$

O seu centro de massa é o ponto $(\bar{x}, \bar{y}, \bar{z})$ com coordenadas definidas por

$$\bar{x}m = \iint_S xf(x, y, z) dS, \quad \bar{y}m = \iint_S yf(x, y, z) dS, \quad \bar{z}m = \iint_S zf(x, y, z) dS.$$

O momento de inércia I_L de S em relação a L é definido por

$$I_L = \iint_S \delta^2(x, y, z) f(x, y, z) dS,$$

representando $\delta(x, y, z)$ a distância a L do ponto genérico da superfície. Como exemplo, determinemos o centro de massa da superfície de uma semi-esfera homogênea de raio a . Consideremos a representação paramétrica

$$\mathbf{r}(u, v) = a \cos u \cos v \mathbf{i} + a \sin u \cos v \mathbf{j} + a \sin v \mathbf{k},$$

com $(u, v) \in [0, 2\pi] \times [0, \pi/2]$. Esta representação particular já foi analisada no Exemplo 2 da Secção 12.2, onde verificámos que a grandeza do produto vectorial fundamental é $a^2 |\cos v|$. Neste exemplo a densidade f é constante, $f = c$, e a massa m e $2\pi a^2 c$, produto da área S por c . Devido à simetria, as coordenadas \bar{x} e \bar{y} do centro de massa são nulas. A cota \bar{z} é dada por

$$\begin{aligned} \bar{z}m &= c \iint_S z dS = c \iint_T a \sin v \cdot a^2 |\cos v| du dv \\ &= 2\pi a^3 c \int_0^{\pi/2} \sin v \cos v dv = \pi a^3 c = \frac{a}{2} m, \end{aligned}$$

pelo que $\bar{z} = a/2$.

EXEMPLO 3. Fluxo de fluido através de uma superfície. Consideremos um fluido como um conjunto de pontos ditos *partículas*. A cada partícula (x, y, z) atribuímos um vector $\mathbf{V}(x, y, z)$ que representa a sua velocidade. Constroi-se assim o campo de velocidade da corrente do fluido. O campo de velocidades pode ou não variar com o tempo. Consideremos unicamente fluxos *estacionários*, isto é, fluxos para os quais a velocidade $\mathbf{V}(x, y, z)$ depende unicamente da posição da partícula e não do tempo.

Designemos por $\rho(x, y, z)$ a densidade (massa por unidade de volume) do fluido no ponto (x, y, z) . Se o fluido é incompressível, a densidade ρ será constante. Se é compressível, tal como um gás, a densidade pode variar de ponto para ponto. Em qualquer das hipóteses a densidade é um campo escalar associado à corrente de fluido. O produto da densidade pela velocidade representamo-lo por \mathbf{F} ; isto é,

$$\mathbf{F}(x, y, z) = \rho(x, y, z) \mathbf{V}(x, y, z).$$

Este é um campo vectorial chamado a *densidade de fluxo* da corrente. O vector $\mathbf{F}(x, y, z)$ tem a mesma direcção que a velocidade, e a sua grandeza tem as dimensões

$$\frac{\text{massa}}{\text{unidade do volume}} \cdot \frac{\text{distância}}{\text{unidade de tempo}} = \frac{\text{massa}}{(\text{unidade de área})(\text{unidade de tempo})}$$

Por outras palavras, o vector *densidade de fluxo* $\mathbf{F}(x, y, z)$ diz-nos quanta massa de fluido, por unidade de área, passa pelo ponto (x, y, z) na direcção de $\mathbf{V}(x, y, z)$ por unidade de tempo.

Seja $S = \mathbf{r}(T)$ uma superfície paramétrica simples. Em cada ponto regular de S designemos por \mathbf{n} o vector unitário normal com sentido concordante com o do produto vectorial fundamental, isto é, seja

$$(12.15) \quad \mathbf{n} = \frac{\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v}}{\left\| \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right\|}.$$

O produto escalar $\mathbf{F} \cdot \mathbf{n}$ representa a componente do vector densidade de fluxo segundo \mathbf{n} . A massa de fluido que passa através de S , na unidade de tempo, na direcção e sentido de \mathbf{n} , define-se pelo integral de superfície

$$\iint_{\mathbf{r}(T)} \mathbf{F} \cdot \mathbf{n} \, dS = \iint_T \mathbf{F} \cdot \mathbf{n} \left\| \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right\| du \, dv.$$

12.8. Mudança de representação paramétrica

Debruçamo-nos agora sobre a análise da independência dos integrais de superfície em face de uma mudança da representação paramétrica daquela. Suponhamos que uma função \mathbf{r} aplica uma região A do plano OUV sobre uma superfície paramétrica $\mathbf{r}(A)$. Supongamos também que A é a imagem da região B no plano st através de uma aplicação biunívoca, continuamente diferenciável, G definida por

$$(12.16) \quad \mathbf{G}(s, t) = U(s, t)\mathbf{i} + V(s, t)\mathbf{j} \quad \text{se } (s, t) \in B.$$

Consideremos a função \mathbf{R} definida em B pela equação

$$(12.17) \quad \mathbf{R}(s, t) = \mathbf{r}[\mathbf{G}(s, t)].$$

(Ver figura 12.11). Duas funções \mathbf{r} e \mathbf{R} assim relacionadas dizem-se *equivalentes regularmente*. Funções equivalentes regularmente representam a mesma superfície, isto é, $\mathbf{r}(A)$ e $\mathbf{R}(B)$ como conjuntos de pontos são idênticos. (Isto é uma consequência do facto de G ser biunívoca.) O teorema que apresentamos a seguir trata da relação entre os respectivos produtos vectoriais fundamentais.

TEOREMA 12.1. *Sejam \mathbf{r} e \mathbf{R} funções equivalentes regularmente relacionadas pela equação (12.17), onde $\mathbf{G} = U\mathbf{i} + V\mathbf{j}$ é uma aplicação biunívoca, continuamente diferenciável, de uma região B do plano OST sobre uma região A do plano OUV , dada por (12.16). Então tem-se*

$$(12.18) \quad \frac{\partial \mathbf{R}}{\partial s} \times \frac{\partial \mathbf{R}}{\partial t} = \left(\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right) \frac{\partial(U, V)}{\partial(s, t)},$$

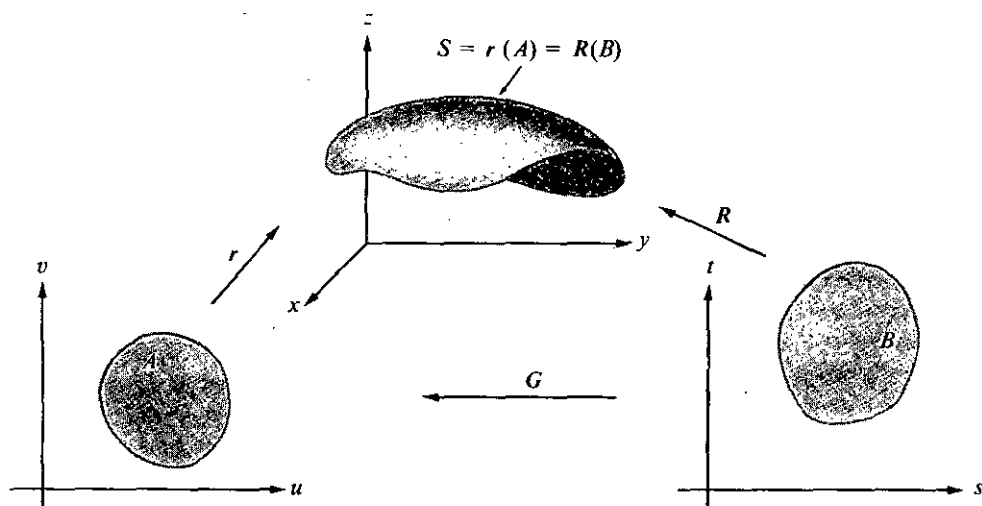


FIG. 12.11. Duas representações paramétricas da mesma superfície.

com as derivadas parciais $\partial \mathbf{r} / \partial u$ e $\partial \mathbf{r} / \partial v$ calculadas no ponto $(U(s, t), V(s, t))$. Quer dizer o produto vectorial fundamental de \mathbf{R} é igual ao de \mathbf{r} multiplicado pelo determinante jacobiano da aplicação G .

Demonstração. As derivadas $\partial \mathbf{R} / \partial s$ e $\partial \mathbf{R} / \partial t$ podem calcular-se por derivação de (12.17). Aplicando a cada componente de \mathbf{R} o teorema 8.8 e recordando os termos, encontramos

$$\frac{\partial \mathbf{R}}{\partial s} = \frac{\partial \mathbf{r}}{\partial u} \frac{\partial U}{\partial s} + \frac{\partial \mathbf{r}}{\partial v} \frac{\partial V}{\partial s} \quad \text{e} \quad \frac{\partial \mathbf{R}}{\partial t} = \frac{\partial \mathbf{r}}{\partial u} \frac{\partial U}{\partial t} + \frac{\partial \mathbf{r}}{\partial v} \frac{\partial V}{\partial t},$$

onde as derivadas $\partial \mathbf{r} / \partial u$ e $\partial \mathbf{r} / \partial v$ são calculadas em $(U(s, t), V(s, t))$. Multiplicando vectorialmente, membro a membro, estas duas equações, tendo presente a ordem dos factores, obtemos

$$\frac{\partial \mathbf{R}}{\partial s} \times \frac{\partial \mathbf{R}}{\partial t} = \left(\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right) \left(\frac{\partial U}{\partial s} \frac{\partial V}{\partial t} - \frac{\partial U}{\partial t} \frac{\partial V}{\partial s} \right) = \left(\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right) \frac{\partial(U, V)}{\partial(s, t)},$$

como queríamos demonstrar.

A invariância dos integrais de superfície para representações paramétricas equivalentes regularmente é agora uma consequência imediata do teorema 12.1.

TEOREMA 12.2. *Sejam r e R funções equivalentes regularmente, como as descritas no teorema 12.1. Se o integral de superfície $\iint_{r(A)} f dS$ existe, então existe também o integral de superfície $\iint_{R(B)} f dS$ e tem-se*

$$\iint_{r(A)} f dS = \iint_{R(B)} f dS.$$

Demonstração. Por definição de integral de superfície temos

$$\iint_{r(A)} f dS = \iint_A f[r(u, v)] \left\| \frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v} \right\| du dv.$$

Usamos em seguida a aplicação G do teorema 12.1 para transformar este integral num integral duplo estendido a uma região B do plano Ost . A fórmula de mudança de variáveis nos integrais duplos garante-nos que

$$\iint_A f[r(u, v)] \left\| \frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v} \right\| du dv = \iint_B f\{r[G(s, t)]\} \left\| \frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v} \right\| \left| \frac{\partial(U, V)}{\partial(s, t)} \right| ds dt,$$

em que as derivadas $\partial r / \partial u$ e $\partial r / \partial v$ do segundo membro são calculadas em $(U(s, t), V(s, t))$. Por causa da equação (12.18), o integral sobre B é igual a

$$\iint_B f[R(s, t)] \left\| \frac{\partial R}{\partial s} \times \frac{\partial R}{\partial t} \right\| ds dt.$$

Mas esta é, por sua vez, a definição do integral de superfície $\iint_{R(B)} f dS$, e a demonstração fica assim completa.

12.9. Outras notações para os integrais de superfície

Se $S = r(T)$ é uma superfície na representação paramétrica, o produto vectorial fundamental $N = \partial r / \partial u \times \partial r / \partial v$ é normal a S em cada ponto regular da superfície. Em cada ponto existem duas normais unitárias, uma n_1 com o sentido de N e outra n_2 com o sentido contrário. Assim,

$$n_1 = \frac{N}{\|N\|} \quad \text{e} \quad n_2 = -n_1.$$

Seja n uma das duas normais n_1 e n_2 . Seja F um campo vectorial definido em S e admitamos que o integral de superfície $\iint_S F \cdot n dS$ existe. Então podemos escrever

$$\begin{aligned}
 (12.19) \quad \iint_S \mathbf{F} \cdot \mathbf{n} \, dS &= \iint_T \mathbf{F}[\mathbf{r}(u, v)] \cdot \mathbf{n}(u, v) \left\| \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right\| du \, dv \\
 &= \pm \iint_T \mathbf{F}[\mathbf{r}(u, v)] \cdot \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} du \, dv,
 \end{aligned}$$

onde se usa o sinal + se $\mathbf{n} = \mathbf{n}_1$ e o sinal - se $\mathbf{n} = \mathbf{n}_2$.

Suponhamos agora que exprimimos \mathbf{F} e \mathbf{r} em função das respectivas componentes, seja

$$\mathbf{F}(x, y, z) = P(x, y, z)\mathbf{i} + Q(x, y, z)\mathbf{j} + R(x, y, z)\mathbf{k}$$

e

$$\mathbf{r}(u, v) = X(u, v)\mathbf{i} + Y(u, v)\mathbf{j} + Z(u, v)\mathbf{k}.$$

Então o produto vectorial fundamental de \mathbf{r} é dado por

$$\mathbf{N} = \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} = \frac{\partial(Y, Z)}{\partial(u, v)} \mathbf{i} + \frac{\partial(Z, X)}{\partial(u, v)} \mathbf{j} + \frac{\partial(X, Y)}{\partial(u, v)} \mathbf{k}.$$

Se $\mathbf{n} = \mathbf{n}_1$, a equação (12.19) vem

$$\begin{aligned}
 (12.20) \quad \iint_S \mathbf{F} \cdot \mathbf{n} \, dS &= \iint_T P[\mathbf{r}(u, v)] \frac{\partial(Y, Z)}{\partial(u, v)} du \, dv \\
 &\quad + \iint_T Q[\mathbf{r}(u, v)] \frac{\partial(Z, X)}{\partial(u, v)} du \, dv + \iint_T R[\mathbf{r}(u, v)] \frac{\partial(X, Y)}{\partial(u, v)} du \, dv;
 \end{aligned}$$

se $\mathbf{n} = \mathbf{n}_2$, cada integral duplo do segundo membro deve ser substituído pelo simétrico. À soma dos integrais duplos do segundo membro dá-se, muitas vezes, a forma mais resumida

$$(12.21) \quad \iint_S P(x, y, z) \, dy \wedge dz + \iint_S Q(x, y, z) \, dz \wedge dx + \iint_S R(x, y, z) \, dx \wedge dy,$$

ou ainda mais abreviadamente

$$(12.22) \quad \iint_S P \, dy \wedge dz + Q \, dz \wedge dx + R \, dx \wedge dy.$$

Os integrais que aparecem em (12.21) e (12.22) são também referidos como integrais de superfície. Assim, por exemplo, o integral de superfície $\iint_S P \, dy \wedge dz$ é definido por

$$(12.23) \quad \iint_S P \, dy \wedge dz = \iint_T P[\mathbf{r}(u, v)] \frac{\partial(Y, Z)}{\partial(u, v)} du \, dv.$$

Esta notação é sugerida pela fórmula de mudança de variáveis num integral duplo.

A despeito da semelhança na notação, o integral do primeiro membro de (12.23) não é um integral duplo. Em primeiro lugar, P é uma função de três variáveis. Em seguida, devemos ter em consideração a ordem segundo a qual os símbolos dy e dz aparecem no integral de superfície, porque

$$\frac{\partial(Y, Z)}{\partial(u, v)} = - \frac{\partial(Z, Y)}{\partial(u, v)}$$

e portanto

$$\iint_S P \, dy \wedge dz = - \iint_S P \, dz \wedge dy.$$

Nesta notação, a fórmula (12.20) vem

$$(12.24) \quad \iint_S \mathbf{F} \cdot \mathbf{n} \, dS = \iint_S P \, dy \wedge dz + Q \, dz \wedge dx + R \, dx \wedge dy$$

se $\mathbf{n} = \mathbf{n}_1$. Se $\mathbf{n} = \mathbf{n}_2$ o integral do segundo membro deve substituir-se pelo simétrico. Esta fórmula é semelhante à seguinte fórmula para os integrais de linha

$$\int_C \mathbf{F} \cdot d\boldsymbol{\alpha} = \int_C P \, dx + Q \, dy + R \, dz.$$

Se o vector unitário normal \mathbf{n} é expresso em função dos cossenos directores, seja

$$\mathbf{n} = \cos \alpha \, \mathbf{i} + \cos \beta \, \mathbf{j} + \cos \gamma \, \mathbf{k},$$

então $\mathbf{F} \cdot \mathbf{n} = P \cos \alpha + Q \cos \beta + R \cos \gamma$, e podemos escrever

$$\iint_S \mathbf{F} \cdot \mathbf{n} \, dS = \iint_S (P \cos \alpha + Q \cos \beta + R \cos \gamma) \, dS.$$

Esta equação é válida quando \mathbf{n} é quer \mathbf{n}_1 , quer \mathbf{n}_2 . Os cossenos directores dependerão da escolha da normal. Se $\mathbf{n} = \mathbf{n}_1$ podemos usar (12.24) para escrevermos

(12.25)

$$\iint_S (P \cos \alpha + Q \cos \beta + R \cos \gamma) \, dS = \iint_S P \, dy \wedge dz + Q \, dz \wedge dx + R \, dx \wedge dy.$$

Se $\mathbf{n} = \mathbf{n}_2$ temos, por sua vez,

(12.26)

$$\iint_S (P \cos \alpha + Q \cos \beta + R \cos \gamma) dS = - \iint_S P dy \wedge dz + Q dz \wedge dx + R dx \wedge dy.$$

12.10. Exercícios

- Seja S a semi-esfera $x^2 + y^2 + z^2 = 1, z \geq 0$ e $F(x, y, z) = x\mathbf{i} + y\mathbf{j}$. Seja \mathbf{n} o vector normal unitário exterior a S . Calcular o valor do integral de superfície $\iint_S \mathbf{F} \cdot \mathbf{n} dS$, utilizando:
 - a representação vectorial $\mathbf{r}(u, v) = \sin u \cos v \mathbf{i} + \sin u \sin v \mathbf{j} + \cos u \mathbf{k}$,
 - a representação explícita $z = \sqrt{1 - x^2 - y^2}$,
- Mostrar que o momento de inércia de uma superfície esférica homogênea em relação a um dos seus diâmetros é igual a $\frac{1}{2}ma^2$, sendo m a massa e a o seu raio.
- Determinar o centro de massa da superfície da semi-esfera homogênea $x^2 + y^2 + z^2 = a^2$, situada acima de OXY .
- Seja S a superfície plana cuja fronteira é o triângulo com vértices $(1, 0, 0)$, $(0, 1, 0)$ e $(0, 0, 1)$, e seja $F(x, y, z) = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$. Represente \mathbf{n} o vector unitário normal a S cuja componente segundo OZ é não negativa. Calcular o integral de superfície $\iint_S \mathbf{F} \cdot \mathbf{n} dS$, utilizando:
 - a representação vectorial $\mathbf{r}(u, v) = (u + v)\mathbf{i} + (u - v)\mathbf{j} + (1 - 2u)\mathbf{k}$,
 - uma representação explícita da forma $z = f(x, y)$
- Seja S uma superfície paramétrica descrita pela fórmula explícita $z = f(x, y)$, com (x, y) a variar numa região plana T , projecção de S no plano XOY . Sejam $\mathbf{F} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$ e \mathbf{n} a normal unitária a S cuja componente segundo OZ é não negativa. Utilizar a representação paramétrica $\mathbf{r}(x, y) = x\mathbf{i} + y\mathbf{j} + f(x, y)\mathbf{k}$ e provar que

$$\iint_S \mathbf{F} \cdot \mathbf{n} dS = \iint_T \left(-P \frac{\partial f}{\partial x} - Q \frac{\partial f}{\partial y} + R \right) dx dy,$$

onde P, Q e R estão calculados em $(x, y, f(x, y))$.

- Seja S a mesma superfície do Exercício 5 e φ um campo escalar. Provar que:

$$(a) \iint_S \varphi(x, y, z) dS = \iint_T \varphi[x, y, f(x, y)] \sqrt{1 + \left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2} dx dy.$$

$$(b) \iint_S \varphi(x, y, z) dy \wedge dz = - \iint_T \varphi[x, y, f(x, y)] \frac{\partial f}{\partial x} dx dy.$$

$$(c) \iint_S \varphi(x, y, z) dz \wedge dx = - \iint_T \varphi[x, y, f(x, y)] \frac{\partial f}{\partial y} dx dy.$$

- Se S é a superfície da esfera $x^2 + y^2 + z^2 = a^2$, calcular o valor do integral de superfície

$$\iint_S xz \, dy \wedge dz + yz \, dz \wedge dx + x^2 \, dx \wedge dy.$$

Escolher uma representação para a qual o produto vectorial fundamental tenha a direcção e o sentido da normal exterior.

8. O cilindro $x^2 + y^2 = 2x$ intersecta uma porção de superfície S na folha superior do cone $x^2 + y^2 = z^2$. Calcular o valor do integral de superfície

$$\iint_S (x^4 - y^4 + y^2 z^2 - z^2 x^2 + 1) \, dS.$$

9. Uma superfície esférica homogênea de raio a é intersectada pela folha superior do cone circular recto cujo vértice está situado no centro da esfera. Se a semi-abertura do cone é α , com $0 < \alpha < \pi$, determine (em função de a e α) o centro de massa da porção de superfície esférica que fica interior ao cone.
10. Uma folha de papel de forma rectangular e homogênea de base $2\pi a$ e altura h é enrolada de modo a dar-lhe a forma de um cilindro circular recto de raio da base a . Calcular o momento de inércia de S em relação a um eixo contendo uma diâmetro da base.
11. Considerar o Exercício 10. Calcular o momento de inércia de S em relação a um eixo situado no plano da base e que é tangente à circunferência que a limita.
12. Uma corrente de fluido tem um vector densidade de fluxo $\mathbf{F}(x, y, z) = x\mathbf{i} - (2x + y)\mathbf{j} + z\mathbf{k}$. Seja S a semi-esfera $x^2 + y^2 + z^2 = 1$, $z \geq 0$, e seja \mathbf{n} o vector unitário normal dirigido para o exterior da esfera. Calcular a massa do fluido passando através de S na unidade de tempo e na direcção de \mathbf{n} .
13. Resolver o Exercício 12 se S inclui também a base plana da semi-esfera. Na base inferior a normal unitária é $-\mathbf{k}$.
14. Seja S a porção do plano $x + y + z = t$ determinada na esfera de raio unidade $x^2 + y^2 + z^2 = 1$. Seja $\varphi(x, y, z) = 1 - x^2 - y^2 - z^2$ se (x, y, z) é interior a esta esfera, e seja $\varphi(x, y, z) = 0$ em qualquer outra hipótese. Provar que

$$\iint_S \varphi(x, y, z) \, dS = \begin{cases} \frac{\pi}{18} (3 - t^2)^2 & \text{se } |t| \leq \sqrt{3}, \\ 0 & \text{se } |t| > \sqrt{3}. \end{cases}$$

[Sugestão: Introduzir novas coordenadas (x_1, y_1, z_1) com o eixo OZ_1 normal ao plano $x + y + z = t$. Usar depois coordenadas polares no plano OX_1Y_1 como parâmetros para S].

12.11. O teorema de Stokes

A parte restante deste capítulo é dedicada especialmente a duas generalizações do segundo teorema fundamental do cálculo aos integrais de superfície. Elas são conhecidas, respectivamente, pelas designações de *teorema de Stokes*† e *teorema da divergência*. Esta Secção trata do primeiro; o teorema da divergência será estudado na Secção 12.19.

† Em honra de G. G. Stokes (1819-1903), matemático irlandês que trouxe contribuições fundamentais a Hidrodinâmica e à óptica.

O teorema de Stokes é uma extensão directa do teorema de Green, o qual estabelece que

$$\iint_S \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx dy = \int_C P dx + Q dy,$$

com S uma região plana limitada por uma curva fechada simples C percorrida no sentido directo. O teorema de Stokes relaciona um integral de superfície com um integral de linha e pode enunciar-se do modo seguinte:

TEOREMA 12.3. TEOREMA DE STOKES. *Suponha-se S uma superfície paramétrica simples regular, $S = \mathbf{r}(T)$, sendo T uma região do plano OUV limitada por uma curva de Jordan Γ , seccionalmente regular. (Ver figura 12.12.) Admita-se também que \mathbf{r} é uma aplicação biunívoca cujas componentes admitem derivadas parciais de segunda ordem contínuas em certo conjunto aberto contendo $T \cup \Gamma$. Seja C a imagem de Γ por \mathbf{r} , e representem P , Q e R campos escalares continuamente diferenciáveis em S . Então tem-se*

$$(12.27) \quad \iint_S \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z} \right) dy \wedge dz + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x} \right) dz \wedge dx + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx \wedge dy \\ = \int_C P dx + Q dy + R dz.$$

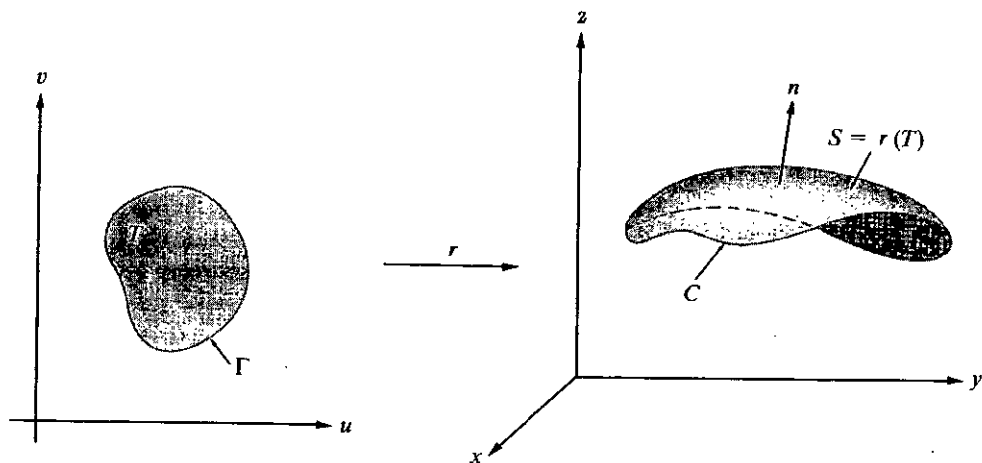


FIG. 12.12. Um exemplo de uma superfície à qual o teorema de Stokes é aplicável.

A curva Γ é descrita no sentido positivo e a curva C no sentido que resulte de aplicar a \mathbf{r} a função \mathbf{r} .

Demonstração: Para demonstrar o teorema basta estabelecer as três seguintes fórmulas,

$$(12.28) \quad \begin{aligned} \int_C P \, dx &= \iint_S \left(-\frac{\partial P}{\partial y} \, dx \wedge dy + \frac{\partial P}{\partial z} \, dz \wedge dx \right), \\ \int_C Q \, dy &= \iint_S \left(-\frac{\partial Q}{\partial z} \, dy \wedge dz + \frac{\partial Q}{\partial x} \, dx \wedge dy \right), \\ \int_C R \, dz &= \iint_S \left(-\frac{\partial R}{\partial x} \, dz \wedge dx + \frac{\partial R}{\partial y} \, dy \wedge dz \right). \end{aligned}$$

A sua adição, membro a membro, dá-nos a fórmula (12.27) do teorema de Stokes. Uma vez que as três são semelhantes, provamos unicamente a primeira.

O caminho para a demonstração consiste em exprimir o integral de superfície como um integral duplo sobre T . Aplica-se então o teorema de Green para exprimir o integral duplo sobre T como um integral de linha ao longo de Γ . Finalmente, mostramos que este integral de linha é igual a $\int_C P \, dx$.

Escrevamos

$$r(u, v) = X(u, v)i + Y(u, v)j + Z(u, v)k$$

e exprimamos o integral de superfície estendido a S na forma

$$\iint_S \left(-\frac{\partial P}{\partial y} \, dx \wedge dy + \frac{\partial P}{\partial z} \, dz \wedge dx \right) = \iint_T \left\{ -\frac{\partial P}{\partial y} \frac{\partial(X, Y)}{\partial(u, v)} + \frac{\partial P}{\partial z} \frac{\partial(Z, X)}{\partial(u, v)} \right\} du \, dv.$$

Designemos por p a função composta dada por

$$p(u, v) = P[X(u, v), Y(u, v), Z(u, v)].$$

O último integrando pode então escrever-se

$$(12.29) \quad -\frac{\partial P}{\partial y} \frac{\partial(X, Y)}{\partial(u, v)} + \frac{\partial P}{\partial z} \frac{\partial(Z, X)}{\partial(u, v)} = \frac{\partial}{\partial u} \left(p \frac{\partial X}{\partial v} \right) - \frac{\partial}{\partial v} \left(p \frac{\partial X}{\partial u} \right).$$

No Exercício 13 da Secção 12.13 esboça-se a demonstração de (12.29). Aplicando o teorema de Green ao integral duplo sobre T obtemos

$$\iint_T \left\{ \frac{\partial}{\partial u} \left(p \frac{\partial X}{\partial v} \right) - \frac{\partial}{\partial v} \left(p \frac{\partial X}{\partial u} \right) \right\} du \, dv = \int_{\Gamma} p \frac{\partial X}{\partial u} du + p \frac{\partial X}{\partial v} dv,$$

sendo Γ percorrida no sentido fronteiro. Parametremos Γ por intermédio de γ definida num intervalo $[a, b]$ e seja

$$(12.30) \quad \alpha(t) = r[\gamma(t)]$$

a correspondente parametrização de C . Expressando então cada integral de linha em termos da sua representação paramétrica encontramos

$$\int_{\Gamma} p \frac{\partial X}{\partial u} du + p \frac{\partial X}{\partial v} dv = \int_C P dx,$$

o que completa a demonstração de (12.28).

12.12. O rotacional e a divergência de um campo vectorial

O integral de superfície que aparece no teorema de Stokes pode exprimir-se de uma forma mais simples por intermédio do *rotacional* de um campo vectorial. Seja F um campo vectorial diferenciável dado por

$$F(x, y, z) = P(x, y, z)\mathbf{i} + Q(x, y, z)\mathbf{j} + R(x, y, z)\mathbf{k}.$$

O rotacional de F é outro campo vectorial definido por

$$(12.31) \quad \text{rot } F = \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z} \right) \mathbf{i} + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x} \right) \mathbf{j} + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) \mathbf{k}.$$

As componentes do $\text{rot } F$ são as funções figurando no integral de superfície da fórmula de Stokes (12.27). Deste modo, o integral de superfície pode escrever-se

$$\iint_S (\text{rot } F) \cdot \mathbf{n} dS,$$

onde \mathbf{n} é o vector normal unitária com o mesmo sentido que o produto vectorial fundamental da superfície, isto é,

$$\mathbf{n} = \frac{\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v}}{\left\| \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right\|}.$$

O integral de linha na fórmula de Stokes (12.27) pode escrever-se na forma $\int_C F \cdot d\alpha$, com α definindo C dada por (12.30). Deste modo, a forma mais simplês para o teorema de Stokes é

$$\iint_S (\text{rot } F) \cdot \mathbf{n} dS = \int_C F \cdot d\alpha.$$

Para o caso especial em que S é a região do plano XOY e $\mathbf{n} = \mathbf{k}$, esta fórmula reduz-se ao teorema de Green

$$\iint_S \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx dy = \int_C P dx + Q dy.$$

A equação (12.31) definindo o rotacional pode memorizar-se facilmente considerando-a como o desenvolvimento de um determinante de terceira ordem,

$$\text{rot } \mathbf{F} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ P & Q & R \end{vmatrix} = \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z} \right) \mathbf{i} + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x} \right) \mathbf{j} + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) \mathbf{k}.$$

Este determinante desenvolve-se segundo os elementos da primeira linha, mas cada "produto" tal como $\partial/\partial y$ vezes R deve interpretar-se como a derivada parcial $\partial R/\partial y$. Podemos, do mesmo modo, escrever esta fórmula como um produto vectorial,

$$\text{rot } \mathbf{F} = \nabla \times \mathbf{F},$$

onde o símbolo ∇ é tratado como se fosse um vector,

$$\nabla = \frac{\partial}{\partial x} \mathbf{i} + \frac{\partial}{\partial y} \mathbf{j} + \frac{\partial}{\partial z} \mathbf{k}.$$

Se formarmos o "produto escalar" $\nabla \cdot \mathbf{F}$ de modo puramente formal, interpretando mais uma vez os produtos tais como $\partial/\partial x$ vezes P como sendo $\partial P/\partial x$, encontramos que

$$(12.32) \quad \nabla \cdot \mathbf{F} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}.$$

A equação (12.32) define um campo escalar chamado a *divergência* de \mathbf{F} , e escreve-se $\text{div } \mathbf{F}$.

Já usamos o símbolo $\nabla \phi$ para representar o gradiente de um campo escalar ϕ , definindo por

$$\nabla \phi = \frac{\partial \phi}{\partial x} \mathbf{i} + \frac{\partial \phi}{\partial y} \mathbf{j} + \frac{\partial \phi}{\partial z} \mathbf{k}.$$

Esta fórmula pode ser interpretada como uma multiplicação formal do símbolo vectorial ∇ pelo campo escalar ϕ . Assim, o gradiente, a divergência, e o rotacional podem representar-se simbolicamente pelos produtos $\nabla \phi$, $\nabla \cdot \mathbf{F}$, $\nabla \times \mathbf{F}$, respectivamente.

Alguns dos teoremas atrás demonstrados podem exprimir-se em termos do rotacional. Por exemplo no teorema (10.9) provámos que um campo vectorial $f = (f_1, \dots, f_n)$, continuamente diferenciável sobre um conjunto conexo aberto S no n -espaço, é um gradiente em S se e só se as derivadas parciais de todas as componentes de f verificam as relações

$$(12.33) \quad D_k f_j(x) = D_j f_k(x) \quad (j, k = 1, 2, \dots, n).$$

Na hipótese do espaço tridimensional o teorema 10.9 pode enunciar-se do modo seguinte:

TEOREMA 12.4. *Se $F = Pi + Qj + Rk$ é um campo vectorial continuamente diferenciável num conjunto conexo S do 3-espaço, então F é um gradiente em S se e só se*

$$(12.34) \quad \text{rot } F = 0 \text{ em } S.$$

Demonstração: No caso tridimensional as relações (12.33) são equivalentes à afirmação de que $\text{rot } F = 0$.

12.13. Exercícios

Em cada um dos Exercícios 1 a 4, transformar o integral de superfície $\iint_S (\text{rot } F) \cdot n \, dS$ em um integral de linha recorrendo ao teorema de Stokes e calcular depois o integral.

1. $F(x, y, z) = y^2i + xyj + xzk$, onde S é a semi-esfera $x^2 + y^2 + z^2 = 1$, $z \geq 0$, e n é a normal unitária com componente segundo OZ não negativa.
2. $F(x, y, z) = yi + zj + xk$, onde S é a porção do paraboloide $z = 1 - x^2 - y^2$ com $z \geq 0$, n é a normal unitária com componente segundo OZ não negativa.
3. $F(x, y, z) = (y - z)i + yzj - xzk$, sendo S formada pelas faces do cubo $0 \leq x \leq 2$, $0 \leq y \leq 2$, $0 \leq z \leq 2$ não situadas no plano XOY . A normal unitária n é dirigida para o exterior.
4. $F(x, y, z) = xzi - yj + x^2yk$, sendo S formada pelas três faces não no plano YOZ do tetraedro limitado pelos três planos coordenados e o plano $3x + y + 3z = 6$. A normal n é a normal unitária dirigida para o exterior do tetraedro.

Nos Exercícios 5 a 10, usar o teorema de Stokes para provar que os integrais de linha tem os valores dados. Em cada exemplo explicar qual o sentido em que deve percorrer-se C para chegar à solução dada.

5. $\int_C ydx + zdy + xdz = \pi a^2 \sqrt{3}$, com C a curva de intersecção da esfera $x^2 + y^2 + z^2 = a^2$ e o plano $x + y + z = 0$.
6. $\int_C (y + z)dx + (z + x)dy + (x + y)dz = 0$, onde C é a curva de intersecção do cilindro $x^2 + y^2 = 2y$ e o plano $y = z$.
7. $\int_C y^2 dx + xy dy + xz dz = 0$, com C a curva do Exercício 6.
8. $\int_C (y - z)dx + (z - x)dy + (x - y)dz = 2\pi a(a + b)$, sendo C a intersecção do cilindro $x^2 + y^2 = a^2$ e o plano $x/a + z/b = 1$, $a > 0$, $b > 0$.
9. $\int_C (y^2 + z^2)dx + (x^2 + z^2)dy + (x^2 + y^2)dz = 2\pi ab^2$, sendo C a curva de intersecção da semi-esfera $x^2 + y^2 + z^2 = 2ax$, $z > 0$, e o cilindro $x^2 + y^2 = 2bx$, com $0 < b < a$.
10. $\int_C (y^2 - z^2)dx + (z^2 - x^2)dy + (x^2 - y^2)dz = 9a^3/2$, sendo C a curva de intersecção do cubo $0 \leq x \leq a$, $0 \leq y \leq a$, $0 \leq z \leq a$ pelo plano $x + y + z = 3a/2$.

11. Se $\mathbf{r} = xi + yj + zk$ e $Pi + Qj + Rk = a \times \mathbf{r}$, sendo a um vector constante, mostrar que $\int_C Pdx + Qdy + Rdz = 2 \int_S a \cdot n dS$, sendo C a curva que limita uma superfície paramétrica S e n a normal unitária a S conveniente.
12. Seja $\mathbf{F} = Pi + Qj + Rk$, onde $P = -y/(x^2 + y^2)$, $Q = x/(x^2 + y^2)$, $R = z$, e seja D o toro gerado pela rotação do círculo $(x - 2)^2 + z^2 = 1$, $y = 0$, em torno do eixo OZ . Provar que $\text{rot } \mathbf{F} = 0$, mas que $\int_C Pdx + Qdy + Rdz$ não é zero se a curva C for a circunferência $x^2 + y^2 = 4$, $z = 0$.
13. Este exercício esboça uma demonstração de (12.29), usada na demonstração do teorema de Stokes.
- (a) Utilizar a fórmula de derivação do produto para provar que

$$\frac{\partial}{\partial u} \left(p \frac{\partial X}{\partial v} \right) - \frac{\partial}{\partial v} \left(p \frac{\partial X}{\partial u} \right) = \frac{\partial p}{\partial u} \frac{\partial X}{\partial v} - \frac{\partial p}{\partial v} \frac{\partial X}{\partial u}.$$

- (b) Seja agora $p(u, v) = P[X(u, v), Y(u, v), Z(u, v)]$. Calcular $\partial p / \partial u$ e $\partial p / \partial v$ e utilizar a alínea (a) para deduzir a equação (12.29),

$$\frac{\partial}{\partial u} \left(p \frac{\partial X}{\partial v} \right) - \frac{\partial}{\partial v} \left(p \frac{\partial X}{\partial u} \right) = - \frac{\partial P}{\partial y} \frac{\partial(X, Y)}{\partial(u, v)} + \frac{\partial P}{\partial z} \frac{\partial(Z, X)}{\partial(u, v)}.$$

12.14. Outras propriedades do rotacional e da divergência

O rotacional e a divergência de um campo vectorial estão relacionados com a matriz jacobiana. Se $\mathbf{F} = Pi + Qj + Rk$, a matriz jacobiana de \mathbf{F} é

$$DF(x, y, z) = \begin{bmatrix} \frac{\partial P}{\partial x} & \frac{\partial P}{\partial y} & \frac{\partial P}{\partial z} \\ \frac{\partial Q}{\partial x} & \frac{\partial Q}{\partial y} & \frac{\partial Q}{\partial z} \\ \frac{\partial R}{\partial x} & \frac{\partial R}{\partial y} & \frac{\partial R}{\partial z} \end{bmatrix}$$

O traço desta matriz (a soma dos elementos da diagonal principal) é a divergência de \mathbf{F} .

Toda a matriz real A pode escrever-se como uma soma de uma matriz simétrica, $\frac{1}{2}(A + A')$, e uma matriz hemi-simétrica, $\frac{1}{2}(A - A')$. Quando A é a matriz jacobiana DF , a parte hemi-simétrica vem

$$(12.35) \quad \frac{1}{2} \begin{bmatrix} 0 & \frac{\partial P}{\partial y} - \frac{\partial Q}{\partial x} & \frac{\partial P}{\partial z} - \frac{\partial R}{\partial x} \\ \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} & 0 & \frac{\partial Q}{\partial z} - \frac{\partial R}{\partial y} \\ \frac{\partial R}{\partial x} - \frac{\partial P}{\partial z} & \frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z} & 0 \end{bmatrix}.$$

Os elementos não nulos desta matriz são as componentes do $\text{rot } F$ e seus simétricos. Se a matriz jacobiana DF é simétrica, todos os elementos de (12.35) são nulos e $\text{rot } F = \mathbf{0}$.

EXEMPLO 1. Sejam $F(x, y, z) = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$. Então tem-se

$$P(x, y, z) = x, \quad Q(x, y, z) = y, \quad R(x, y, z) = z,$$

e a correspondente matriz jacobiana é a matriz identidade 3×3 . Por conseguinte

$$\text{div } F = 3 \quad \text{e} \quad \text{rot } F = \mathbf{0}.$$

Mais geralmente, se $F(x, y, z) = f(x)\mathbf{i} + g(y)\mathbf{j} + h(z)\mathbf{k}$, a matriz jacobiana tem os elementos $f'(x)$, $g'(y)$, $h'(z)$ na diagonal principal e nulos os restantes elementos, pelo que

$$\text{div } F = f'(x) + g'(y) + h'(z) \quad \text{e} \quad \text{rot } F = \mathbf{0}.$$

EXEMPLO 2. Seja $F(x, y, z) = xy^2z^2\mathbf{i} + z^2 \text{sen } y\mathbf{j} + x^2e^y\mathbf{k}$. A matriz jacobiana é

$$\begin{bmatrix} y^2z^2 & 2xyz^2 & 2xy^2z \\ 0 & z^2 \cos y & 2z \text{sen } y \\ 2xe^y & x^2e^y & 0 \end{bmatrix}.$$

Por conseguinte

$$\text{div } F = y^2z^2 + z^2 \cos y$$

e

$$\text{rot } F = (x^2e^y - 2z \text{sen } y)\mathbf{i} + (2xyz^2 - 2xe^y)\mathbf{j} - 2xyz^2\mathbf{k}.$$

EXEMPLO 3. A divergência e o rotacional de um gradiente. Suponhamos que F é um gradiente, por exemplo $F = \text{grad } \varphi = \partial\varphi/\partial x\mathbf{i} + \partial\varphi/\partial y\mathbf{j} + \partial\varphi/\partial z\mathbf{k}$. A matriz jacobiana é

$$(12.36) \quad \begin{bmatrix} \frac{\partial^2 \varphi}{\partial x^2} & \frac{\partial^2 \varphi}{\partial y \partial x} & \frac{\partial^2 \varphi}{\partial z \partial x} \\ \frac{\partial^2 \varphi}{\partial x \partial y} & \frac{\partial^2 \varphi}{\partial y^2} & \frac{\partial^2 \varphi}{\partial z \partial y} \\ \frac{\partial^2 \varphi}{\partial x \partial z} & \frac{\partial^2 \varphi}{\partial y \partial z} & \frac{\partial^2 \varphi}{\partial z^2} \end{bmatrix}.$$

Deste modo,

$$\text{div } F = \frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial y^2} + \frac{\partial^2 \varphi}{\partial z^2}.$$

A expressão do segundo membro chama-se o *laplaciano* de φ e escreve-se muitas vezes, mas abreviadamente $\nabla^2 \varphi$. Assim, a divergência de um gradiente $\nabla \varphi$ é o laplaciano de φ . Em notação simbólica escreve-se

$$(12.37) \quad \operatorname{div}(\nabla \varphi) = \nabla^2 \varphi.$$

Quando $\nabla^2 \varphi = 0$, a função φ diz-se *harmónica*. A equação (12.37) mostra que o gradiente de uma função harmónica tem divergência nula. Se as derivadas parciais mistas na matriz (12.36) são contínuas, a matriz é simétrica e o $\operatorname{rot} F$ é zero. Por outras palavras,

$$\operatorname{rot}(\operatorname{grad} \varphi) = \mathbf{0}$$

para todo o campo escalar φ com derivadas parciais de segunda ordem mistas contínuas. Este exemplo mostra que a condição $\operatorname{rot} F = \mathbf{0}$ é necessária para que um campo vectorial F , continuamente diferenciável, seja um gradiente. Quer dizer, se $\operatorname{rot} F \neq \mathbf{0}$ num conjunto aberto S , então F não é um gradiente em S . Sabemos também, do teorema 12.4, que se $\operatorname{rot} F = \mathbf{0}$ num conjunto *convexo* aberto S , então F é um gradiente em S . Um campo com rotacional nulo diz-se *irrotacional*.

EXEMPLO 4. Um campo vectorial com divergência nula e rotacional nulo. Seja S o conjunto de todos $(x, y) \neq (0, 0)$, e seja

$$F(x, y) = -\frac{y}{x^2 + y^2} \mathbf{i} + \frac{x}{x^2 + y^2} \mathbf{j}$$

se $(x, y) \in S$. Do Exemplo 2 da Secção 10.6 sabemos que F não é um gradiente em S (se bem que F seja um gradiente em todo o rectângulo que não contenha a origem). A matriz jacobiana é

$$DF(x, y) = \begin{bmatrix} \frac{2xy}{(x^2 + y^2)^2} & \frac{y^2 - x^2}{(x^2 + y^2)^2} & 0 \\ \frac{y^2 - x^2}{(x^2 + y^2)^2} & \frac{-2xy}{(x^2 + y^2)^2} & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

onde resulta de imediato que $\operatorname{div} F = 0$ e $\operatorname{rot} F = \mathbf{0}$ em S .

EXEMPLO 5. A divergência e o rotacional de um rotacional. Se $F = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$, o rotacional de F é um novo campo vectorial e é possível calcular a sua divergência e o seu rotacional. A matriz jacobiana do $\operatorname{rot} F$ é

$$\begin{bmatrix} \frac{\partial^2 R}{\partial x \partial y} - \frac{\partial^2 Q}{\partial x \partial z} & \frac{\partial^2 R}{\partial y^2} - \frac{\partial^2 Q}{\partial y \partial z} & \frac{\partial^2 R}{\partial z \partial y} - \frac{\partial^2 Q}{\partial z^2} \\ \frac{\partial^2 P}{\partial x \partial z} - \frac{\partial^2 R}{\partial x^2} & \frac{\partial^2 P}{\partial y \partial z} - \frac{\partial^2 R}{\partial y \partial x} & \frac{\partial^2 P}{\partial z^2} - \frac{\partial^2 R}{\partial z \partial x} \\ \frac{\partial^2 Q}{\partial x^2} - \frac{\partial^2 P}{\partial x \partial y} & \frac{\partial^2 Q}{\partial y \partial x} - \frac{\partial^2 P}{\partial y^2} & \frac{\partial^2 Q}{\partial z \partial x} - \frac{\partial^2 P}{\partial z \partial y} \end{bmatrix}.$$

Se admitimos que todas as derivadas parciais mistas são contínuas, encontramos que

$$\operatorname{div}(\operatorname{rot} F) = 0$$

e

$$(12.38) \quad \operatorname{rot}(\operatorname{rot} F) = \operatorname{grad}(\operatorname{div} F) - \nabla^2 F,$$

sendo $\nabla^2 F$ definido pela equação

$$\nabla^2 F = (\nabla^2 P)i + (\nabla^2 Q)j + (\nabla^2 R)k.$$

A identidade (12.38) relaciona os quatro operadores, gradiente, divergência, rotacional e laplaciano. No Exercício 7 da Secção 12.15 é solicitada a demonstração de (12.38).

O rotacional e a divergência possuem algumas propriedades gerais análogas às da derivação ordinária. Em primeiro lugar, são *operadores lineares*, isto é, se a e b são constantes, então

$$(12.39) \quad \operatorname{div}(aF + bG) = a \operatorname{div} F + b \operatorname{div} G,$$

e

$$(12.40) \quad \operatorname{rot}(aF + bG) = a \operatorname{rot} F + b \operatorname{rot} G.$$

Gozam também da propriedade análoga a regra da derivação para o produto:

$$(12.41) \quad \operatorname{div}(\varphi F) = \varphi \operatorname{div} F + \nabla \varphi \cdot F,$$

e

$$(12.42) \quad \operatorname{rot}(\varphi F) = \varphi \operatorname{rot} F + \nabla \varphi \times F,$$

com φ qualquer campo escalar diferenciável. Estas propriedades são consequências imediatas das definições de rotacional e divergência; no Exercício da Secção 12.15 solicita-se a respectiva demonstração.

Se utilizarmos mais uma vez o vector simbólico

$$\nabla = \frac{\partial}{\partial x} \mathbf{i} + \frac{\partial}{\partial y} \mathbf{j} + \frac{\partial}{\partial z} \mathbf{k}$$

cada uma das fórmulas (12.41) e (12.42) toma uma forma muito mais parecida com a regra usual de derivação do produto:

$$\nabla \cdot (\varphi \mathbf{F}) = \varphi \nabla \cdot \mathbf{F} + \nabla \varphi \cdot \mathbf{F}$$

e

$$\nabla \times (\varphi \mathbf{F}) = \varphi \nabla \times \mathbf{F} + \nabla \varphi \times \mathbf{F}.$$

No Exemplo 3 o laplaciano de um campo escalar, $\nabla^2 \varphi$, definiu-se como sendo $\partial^2 \varphi / \partial x^2 + \partial^2 \varphi / \partial y^2 + \partial^2 \varphi / \partial z^2$. No Exemplo 5 o laplaciano $\nabla^2 \mathbf{F}$ de um campo vectorial definiu-se pelas suas componentes. Obtêm-se fórmulas correctas para $\nabla^2 \varphi - \nabla^2 \mathbf{F}$ se interpretarmos ∇^2 como o operador simbólico

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.$$

Esta fórmula para ∇^2 também se obterá pelo “produto escalar” deste vector simbólico ∇ por si próprio. Então, temos $\nabla^2 = \nabla \cdot \nabla$, e podemos escrever

$$\nabla^2 \varphi = (\nabla \cdot \nabla) \varphi \quad \text{e} \quad \nabla^2 \mathbf{F} = (\nabla \cdot \nabla) \mathbf{F}.$$

Consideremos agora a fórmula $\nabla \cdot \nabla \varphi$. Isto pode interpretar-se como $(\nabla \cdot \nabla) \varphi$ que é $\nabla^2 \varphi$; ou como $\nabla \cdot (\nabla \varphi)$, que é $\text{div}(\nabla \varphi)$. No Exemplo 3 mostramos que $\text{div}(\nabla \varphi) = \nabla^2 \varphi$, pelo que se terá

$$(\nabla \cdot \nabla) \varphi = \nabla \cdot (\nabla \varphi);$$

logo podemos escrever $\nabla \cdot \nabla \varphi$ para uma ou outra dessas expressões sem perigo de ambiguidade. Tal não é, porém, verdadeiro, quando φ é substituído pelo campo vectorial \mathbf{F} . A expressão $(\nabla \cdot \nabla) \mathbf{F}$ é $\nabla^2 \mathbf{F}$, que já se definiu. Todavia, $\nabla \cdot (\nabla \mathbf{F})$ não tem significado porque $\nabla \mathbf{F}$ não está definido. Portanto a expressão $\nabla \cdot \nabla \mathbf{F}$ é provida de significado unicamente quando é interpretada como $(\nabla \cdot \nabla) \mathbf{F}$. Estas observações fazem sobressair o facto de que embora as fórmulas simbólicas apareçam muitas vezes como convenientes no manejo e retenção de memória, é necessária uma utilização cautelosa.

12.15. Exercícios

- Para cada um dos campos vectoriais seguintes determinar a matriz jacobiana e calcular o rotacional e a divergência
 - $\mathbf{F}(x, y, z) = (x^2 + yz)\mathbf{i} + (y^2 + xz)\mathbf{j} + (z^2 + xy)\mathbf{k}$.
 - $\mathbf{F}(x, y, z) = (2z - 3y)\mathbf{i} + (3x - z)\mathbf{j} + (y - 2x)\mathbf{k}$.

- (c) $F(x, y, z) = (z + \sin y)\mathbf{i} - (z - x \cos y)\mathbf{j}$,
 (d) $F(x, y, z) = e^{xy}\mathbf{i} + \cos xy\mathbf{j} + \cos xz^2\mathbf{k}$.
 (e) $F(x, y, z) = x^2 \sin y \mathbf{i} + y^2 \sin xz \mathbf{j} + xy \sin(\cos z) \mathbf{k}$.
- Se $\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ e $r = \|\mathbf{r}\|$, calcular $\text{rot } [f(r)\mathbf{r}]$, sendo f uma função derivável.
 - Se $\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ e \mathbf{A} é um vector constante, mostrar que $\text{rot } (\mathbf{A} \times \mathbf{r}) = 2\mathbf{A}$.
 - Se $\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ e $r = \|\mathbf{r}\|$, determinar todos os inteiros n para os quais $\text{div } (r^n \mathbf{r}) = 0$.
 - Determinar um campo vectorial cujo rotacional é $x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ ou provar que não existe um tal campo vectorial.
 - Demonstrar as propriedades elementares do rotacional e da divergência indicada em (12.39) e (12.42).
 - Demonstrar que $\text{rot } (\text{rot } \mathbf{F}) = \text{grad } (\text{div } \mathbf{F}) - \nabla^2 \mathbf{F}$ na hipótese de serem contínuas as derivadas parciais mistas de segunda ordem das componentes de \mathbf{F} .
 - Provar a identidade

$$\nabla \cdot (\mathbf{F} \times \mathbf{G}) = \mathbf{G} \cdot (\nabla \times \mathbf{F}) - \mathbf{F} \cdot (\nabla \times \mathbf{G}),$$

com \mathbf{F} e \mathbf{G} campos vectoriais diferenciáveis.

- Um campo vectorial \mathbf{F} não será gradiente de um potencial a menos que $\text{rot } \mathbf{F} = \mathbf{0}$. Todavia pode ser possível determinar um campo escalar não nulo μ tal que $\mu\mathbf{F}$ seja um gradiente. Provar que se um tal μ existir, \mathbf{F} é sempre perpendicular ao seu rotacional. Quando o campo é bidimensional, por exemplo $\mathbf{F} = P\mathbf{i} + Q\mathbf{j}$, este exercício constitui uma condição necessária para que a equação diferencial $Pdx + Qdy = 0$ admita um factor integrante. (A inversa é também verdadeira, isto é, se $\mathbf{F} \cdot \text{rot } \mathbf{F} = 0$ numa região conveniente, existe um μ não nulo tal que $\mu\mathbf{F}$ seja um gradiente. A demonstração do recíproco não pedida).
- Seja $\mathbf{F}(x, y, z) = y^2z^2\mathbf{i} + z^2x^2\mathbf{j} + x^2y^2\mathbf{k}$. Mostrar que $\text{rot } \mathbf{F}$ nem sempre é nulo, mas que $\mathbf{F} \cdot \text{rot } \mathbf{F} = 0$. Determinar em campo escalar μ tal que $\mu\mathbf{F}$ é um gradiente.
- Sejam $\mathbf{V}(x, y) = y^c\mathbf{i} + x^c\mathbf{j}$, com c uma constante positiva, e $\mathbf{r}(x, y) = x\mathbf{i} + y\mathbf{j}$. Seja R uma região plana limitada por uma curva de Jordan C , seccionalmente regular. Calcular $\text{div } (\mathbf{V} \times \mathbf{r})$ e $\text{rot } (\mathbf{V} \times \mathbf{r})$, e usar o teorema de Green para provar que

$$\oint_C \mathbf{V} \times \mathbf{r} \cdot d\boldsymbol{\alpha} = 0,$$

sendo $\boldsymbol{\alpha}$ a função que define C .

- Demonstrar que o teorema de Green pode traduzir-se pela igualdade

$$\iint_R (\text{rot } \mathbf{V}) \cdot \mathbf{k} \, dx \, dy = \oint_C \mathbf{V} \cdot \mathbf{T} \, ds,$$

com \mathbf{T} a tangente unitária a C e s o comprimento de arco.

- Uma região plana R é limitada por uma curva de Jordan seccionalmente regular C . Os momentos de inércia de R em relação aos eixos OX e OY são, respectivamente, a e b . Calcular o integral de linha

$$\oint_C \nabla(r^4) \cdot \mathbf{n} \, ds$$

em função de a e b . No integral $r = \|x\mathbf{i} + y\mathbf{j}\|$, \mathbf{n} representa a normal unitária dirigida para o exterior de C , e s representa o comprimento de arco. A curva é descrita no sentido directo.

- Seja \mathbf{F} um campo vectorial bidimensional. Dar uma definição do integral de linha $\int_C \mathbf{F} \times d\boldsymbol{\alpha}$. Essa definição deve ser tal que a fórmula seguinte seja uma consequência do teorema de Green:

$$\int_C \mathbf{F} \times d\alpha = k \iint_R (\operatorname{div} \mathbf{F}) dx dy,$$

com R uma região plana limitada por uma curva fechada simples C .

★ 12.16. Reconstrução de um campo vectorial a partir do seu rotacional

Ao estudarmos o gradiente aprendemos a verificar se um campo vectorial é ou não um gradiente. Consideremos agora uma questão análoga referente ao rotacional. Dado um campo vectorial \mathbf{F} , existirá um \mathbf{G} tal que $\operatorname{rot} \mathbf{G} = \mathbf{F}$? Suponhamos que escrevemos $\mathbf{F} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$ e $\mathbf{G} = L\mathbf{i} + M\mathbf{j} + N\mathbf{k}$. Para resolvermos a equação $\operatorname{rot} \mathbf{G} = \mathbf{F}$ temos que resolver o sistema de equações de derivadas parciais

$$(12.43) \quad \frac{\partial N}{\partial y} - \frac{\partial M}{\partial z} = P, \quad \frac{\partial L}{\partial z} - \frac{\partial N}{\partial x} = Q, \quad \frac{\partial M}{\partial x} - \frac{\partial L}{\partial y} = R$$

para as três funções desconhecidas L , M e N , quando são dados P , Q e R .

Nem sempre é possível resolver tal sistema. Por exemplo, provámos na Secção 12.14 que a divergência de um rotacional é sempre zero. Por conseguinte, para o sistema (12.43) admitir uma solução em algum conjunto aberto S é necessário ter

$$(12.44) \quad \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z} = 0$$

em todo S . Esta condição é também suficiente para que o sistema (12.43) tenha uma solução se restringirmos de maneira conveniente o conjunto S em que se verifica (12.44). Demonstramos a seguir que a condição (12.44) é suficiente, quando S é um intervalo tridimensional.

TEOREMA 12.5. *Se \mathbf{F} é continuamente diferenciável num intervalo aberto S do espaço tridimensional, então existe um campo vectorial \mathbf{G} tal que $\operatorname{rot} \mathbf{G} = \mathbf{F}$ se e só se $\operatorname{div} \mathbf{F} = 0$ em todo S .*

Demonstração. A necessidade da condição $\operatorname{div} \mathbf{F} = 0$ já foi estabelecida, uma vez que a divergência de um rotacional é sempre zero. Para estabelecer a suficiência dessa condição temos que determinar três funções L , M e N que satisfaçam às três equações (12.43). Tentemos resolver o problema partindo da hipótese de que $L = 0$. Então as segunda e terceira equações (12.43) escrevem-se

$$\frac{\partial N}{\partial x} = -Q \quad \text{e} \quad \frac{\partial M}{\partial x} = R.$$

Isto significa que deve ser

$$N(x, y, z) = - \int_{x_0}^x Q(t, y, z) dt + f(y, z)$$

e

$$M(x, y, z) = \int_{x_0}^x R(t, y, z) dt + g(y, z),$$

onde cada integração é feita ao longo de um segmento de recta contido em S e as “constantes de integração” $f(y, z)$ e $g(y, z)$ são independentes de x . Tentemos determinar uma solução com $f(y, z) = 0$. A primeira equação (12.43) exige

$$(12.45) \quad \frac{\partial N}{\partial y} - \frac{\partial M}{\partial z} = P.$$

Para a escolha de M e N acabada de referir vem:

$$(12.46) \quad \frac{\partial N}{\partial y} - \frac{\partial M}{\partial z} = - \frac{\partial}{\partial y} \int_{x_0}^x Q(t, y, z) dt - \frac{\partial}{\partial z} \int_{x_0}^x R(t, y, z) dt - \frac{\partial g}{\partial z}.$$

Permutamos em seguida, entre si, as operações de derivação e integração, aplicando o teorema 10.8. Quer dizer que podemos escrever

$$(12.47) \quad \frac{\partial}{\partial y} \int_{x_0}^x Q(t, y, z) dt = \int_{x_0}^x D_2 Q(t, y, z) dt$$

e

$$(12.48) \quad \frac{\partial}{\partial z} \int_{x_0}^x R(t, y, z) dt = \int_{x_0}^x D_3 R(t, y, z) dt.$$

A equação (12.46) escreve-se então

$$(12.49) \quad \frac{\partial N}{\partial y} - \frac{\partial M}{\partial z} = \int_{x_0}^x [-D_2 Q(t, y, z) - D_3 R(t, y, z)] dt - \frac{\partial g}{\partial z}.$$

Usando a condição (12.44) podemos substituir o integrando em (12.49) por $D_1 P(t, y, z)$; a equação (12.49) vem

$$\frac{\partial N}{\partial y} - \frac{\partial M}{\partial z} = \int_{x_0}^x D_1 P(t, y, z) dt - \frac{\partial g}{\partial z} = P(x, y, z) - P(x_0, y, z) - \frac{\partial g}{\partial z}.$$

Portanto (12.45) será satisfeita se escolhermos g de maneira que $\partial g / \partial z = -P(x_0, y, z)$. Assim, por exemplo, podemos tomar

$$g(y, z) = - \int_{z_0}^z P(x_0, y, u) du.$$

Este argumento leva-nos a considerar o campo vectorial $G = Li + Mj + Nk$, onde $L(x, y, z) = 0$ e

$$M(x, y, z) = \int_{x_0}^x R(t, y, z) dt - \int_{x_0}^z P(x_0, y, u) du, \quad N(x, y, z) = - \int_{x_0}^z Q(t, y, z) dt.$$

Com esta escolha de L , M e N é fácil verificar recorrendo a (12.47) e (12.48), que as três equações (12.43) são satisfeitas, conduzindo-nos a $\text{rot } G = F$ como pretendíamos.

Deve ter-se presente que a precedente demonstração não somente estabelece a existência de um campo vectorial G cujo rotacional é F , mas também fornece um método directo para determinar G por integração respeitante às componentes de F .

Para um dado F , o campo vectorial G que construímos não é a única solução da equação $\text{rot } G = F$. Se adicionarmos a este G qualquer gradiente $\nabla \phi$, continuamente diferenciável, obtemos outra solução porque

$$\text{rot } (G + \nabla \phi) = \text{rot } G + \text{rot } (\nabla \phi) = \text{rot } G = F,$$

uma vez que $\text{rot } (\nabla \phi) = 0$. Além disso, é fácil provar que *todas* as soluções continuamente diferenciáveis devem ser da forma $G + \nabla \phi$. Na realidade, se H é outra solução, então $\text{rot } H = \text{rot } G$, pelo que $\text{rot } (H - G) = 0$. Pelo teorema 10.9 resulta que $H - G = \nabla \phi$ para algum gradiente $\nabla \phi$ continuamente diferenciável; consequentemente $H = G + \nabla \phi$, como afirmámos.

Um campo vectorial F para o qual $\text{div } F = 0$ costuma designar-se por *campo solenoidal*. O teorema 12.5 estabelece que um campo vectorial é solenoidal num intervalo aberto S do espaço tridimensional se e só se for o rotacional de outro campo vectorial em S .

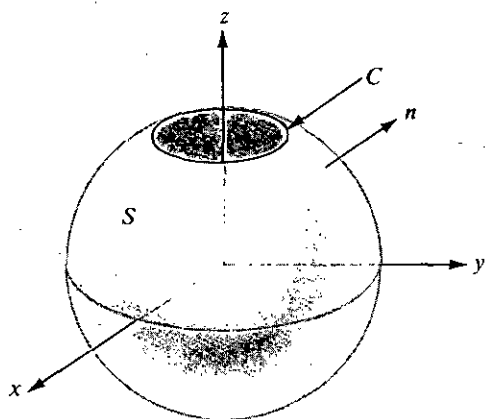
O exemplo que se segue mostra que esta proposição não é verdadeira para conjuntos abertos *quaisquer*.

EXEMPLO. *Um campo vectorial solenoidal que não é um rotacional.* Seja D a região do espaço tridimensional limitado por duas esferas concêntricas com centro na origem a raios a e b , sendo $0 < a < b$. Seja $V = r/r^3$, $r = xi + yj + zk$ e $r = \|r\|$. É fácil verificar que $\text{div } V = 0$ em toda a região D . Com efeito, temos a fórmula geral

$$\text{div } (r^n r) = (n + 3)r^n,$$

sendo neste exemplo $n = -3$. Usaremos o teorema de Stokes para provar que este V não é um rotacional em D (embora o seja em todo o intervalo aberto tridimensional que não contenha a origem). Para isso admitamos que existe um campo vectorial U tal que $V = \text{rot } U$ em D e chegamos a uma contradição. Pelo teorema de Stokes podemos escrever

$$(12.50) \quad \iint_S (\text{rot } U) \cdot n \, dS = \oint_C U \cdot d\alpha,$$

FIG. 12.13. A superfície S e a curva C da equação (12.50).

onde C e S são a curva e a superfície representadas na figura 12.13. Para construir S , tomamos uma superfície esférica de raio R concêntrica com as fronteiras de D , sendo $a < R < b$, e destacamos uma pequena "calote polar", como se indica na figura. A parte restante é a superfície S . A curva C é o bordo circular que se indica. Seja n a normal unitária dirigida para o exterior de S , de maneira que $n = r/r$. Uma vez que $\text{rot } U = V = r/r^3$, temos

$$(\text{rot } U) \cdot n = \frac{r}{r^3} \cdot \frac{r}{r} = \frac{1}{r^2}.$$

Sobre a superfície S este produto escalar tem o valor constante $1/R^2$. Deste modo tem-se

$$\iint_S (\text{rot } U) \cdot n \, dS = \frac{1}{R^2} \iint_S dS = \frac{\text{área de } S}{R^2}.$$

Quando a "calote polar" se reduz a um ponto, a área de S tende para $4\pi R^2$ (a área de toda a esfera) e, por conseguinte, o valor do integral de superfície em (12.50) tende para 4π .

Examinemos em seguida o integral de linha em (12.50). É fácil provar que para qualquer integral de linha $\int_C U \cdot d\alpha$ se verifica a desigualdade

$$\left| \int_C U \cdot d\alpha \right| \leq M \cdot (\text{comprimento de } C),$$

sendo M uma constante dependendo de U . (Com efeito, M pode tomar-se como sendo o máximo valor de $\|U\|$ ao longo de C .) Deste modo, ao reduzir-se a "calote polar" a um ponto, o comprimento de C e o valor do integral de linha tendem para zero. Então

estamos perante uma contradição; o integral de superfície em (12.50) pode fazer-se tomar um valor tão próximo quanto se queira de 4π e o integral de linha, a que é igual, tomar um valor arbitrariamente próximo de zero. Portanto não pode existir na região D uma função U cuja rotacional seja V .

A dificuldade aqui é dividida à estrutura geométrica da região D . Embora tal região seja simplesmente conexa (isto é, qualquer curva fechada simples em D é o bordo de uma superfície paramétrica contida totalmente em D), existem *superfícies* fechadas em D que não são fronteiras completas de sólidos contidos inteiramente em D . Por exemplo, nenhuma esfera centrada na origem é a fronteira completa de um sólido completamente contido em D . Se a região D goza de propriedade de que *cada* superfície fechada em D é a fronteira de um sólido completamente situado em D , pode provar-se que existe um campo vectorial U tal que $V = \text{rot } U$ em D se e só se $\text{div } V = 0$ em toda a região D . A demonstração desta proposição é difícil e não será, por tal motivo, apresentada aqui.

★ 12.17. Exercícios

1. Determinar um campo vectorial $G(x, y, z)$ cujo rotacional é $2i + j + 3k$, em todo E_3 . Qual é o campo vectorial mais geral, continuamente diferenciável, gozando desta propriedade?
2. Provar que o campo vectorial $F(x, y, z) = (y - z)i + (z - x)j + (x - y)k$ é solenoidal, e determinar um campo vectorial G tal que $F = \text{rot } G$ em todo o espaço (tridimensional).
3. Seja $F(x, y, z) = -zi + xyk$. Determinar um campo vectorial continuamente diferenciável G , da forma $G(x, y, z) = L(x, y, z)i + M(x, y, z)j$, tal que $F = \text{rot } G$ em todo o espaço. Qual é a forma mais geral de G ?
4. Se dois campos vectoriais U e V são irrotacionais, mostrar que o campo vectorial $U \times V$ é solenoidal.
5. Seja $r = xi + yj + zk$ e $r = \|r\|$. Mostrar que $n = -3$ é o único valor de n para o qual $r^n r$ é, solenoidal para $r \neq 0$. Para este n , escolher um intervalo tridimensional S que não contenha a origem a exprimir $r^3 r$ como um rotacional em S . *Nota:* Embora $r^3 r$ seja um rotacional num tal S , não é um rotacional no conjunto de todos os pontos distintos de $(0, 0, 0)$.
6. Determinar a forma mais geral da função f de uma variável, continuamente derivável, tal que o campo vectorial $f(r)r$ seja solenoidal, sendo $r = xi + yj + zk$ e $r = \|r\|$.
7. Seja V um campo vectorial continuamente diferenciável em certo intervalo aberto do espaço tridimensional. Considerem-se as duas seguintes proposições relativas a V :
(i) $\text{rot } V = 0$ e $V = \text{rot } U$, para algum campo vectorial U continuamente diferenciável (em todo S).
(ii) Existe um campo escalar ϕ tal que $\nabla\phi$ seja continuamente diferenciável e tal que

$$V = \text{grad } \phi \quad \text{e} \quad \nabla^2 \phi = 0 \quad \text{em todo } S.$$

- (a) Provar que (i) implica (ii). Quer dizer, um campo vectorial que é simultaneamente irrotacional e solenoidal em S é o gradiente de uma função harmónica em S .
 - (b) Provar que (ii) implica (i), ou então apresentar um contra exemplo.
8. Admitir que todos os campos considerados são continuamente diferenciáveis em S (intervalo aberto). Seja $H = F + G$, com F solenoidal e G irrotacional. Então existe um campo vectorial U , tal que $F = \text{rot } U$ e um campo escalar ϕ tal que $G = \nabla\phi$ em S . Provar que U e ϕ satisfazem às seguintes equações de derivadas parciais

$$\nabla^2 \varphi = \operatorname{div} \mathbf{H}, \quad \operatorname{grad} (\operatorname{div} \mathbf{U}) - \nabla^2 \mathbf{U} = \operatorname{rot} \mathbf{H}.$$

Nota: Este exercício tem largas aplicações, porque pode demonstrar-se que todo o campo vectorial \mathbf{H} , continuamente diferenciável em S , pode exprimir-se na forma $\mathbf{H} = \mathbf{F} + \mathbf{G}$, em que \mathbf{F} é solenoidal e \mathbf{G} irrotacional.

9. Seja $\mathbf{H}(x, y, z) = x^2 y \mathbf{i} + y^2 z \mathbf{j} + z^2 x \mathbf{k}$. Determinar os campos vectoriais \mathbf{F} e \mathbf{G} , com \mathbf{F} um rotacional e \mathbf{G} um gradiente, tais que $\mathbf{H} = \mathbf{F} + \mathbf{G}$.
10. Sejam μ e v campos escalares continuamente diferenciáveis num intervalo aberto R do espaço tridimensional.
 - (a) Mostrar que existe um campo vectorial \mathbf{F} tal que $\nabla u \times \nabla v = \operatorname{rot} \mathbf{F}$ em todo R .
 - (b) Determinar se sim ou não qualquer dos três campos vectoriais seguintes podem ser utilizados como \mathbf{F} na alínea (a): (i) $\nabla(uv)$; (ii) $u \nabla v$; (iii) $v \nabla u$.
 - (c) Se $u(x, y, z) = x^3 - y^3 + z^3$ e $v(x, y, z) = x + y + z$, calcular o integral de superfície

$$\iint_S \nabla u \times \nabla v \cdot \mathbf{n} \, dS,$$

sendo S a semi-esfera $x^2 + y^2 + z^2 = 1, z \geq 0$, e \mathbf{n} o vector normal unitário com componente segundo OZ não negativa.

12.18. Extensões do teorema de Stokes

O teorema de Stokes pode generalizar-se a superfícies regulares simples, mas mais gerais. Se T é uma região multiplamente conexa semelhante à que se apresenta na figura 12.14 (com um número finito de buracos), a imagem $S = \mathbf{r}(T)$ conterá o mesmo número de buracos que T . Para generalizar o teorema de Stokes a tais superfícies seguiremos o mesmo tipo de raciocínio que na demonstração precedente, excepto em que usaremos o teorema de Green para regiões multiplamente conexas (teorema

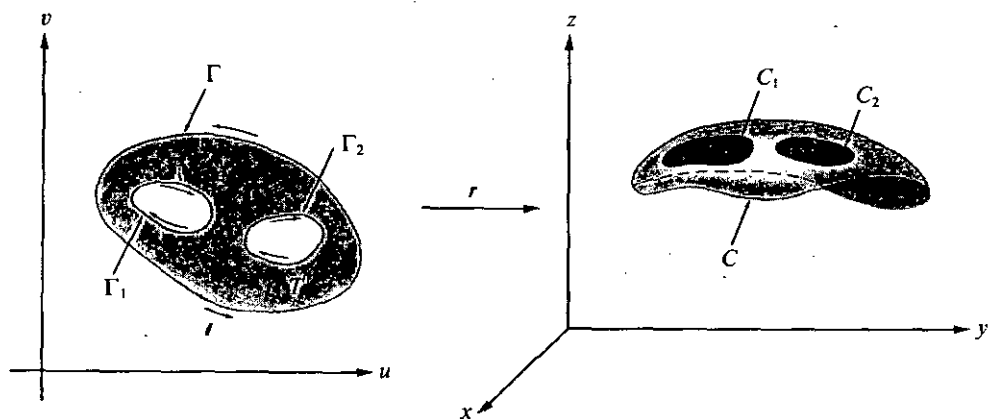


FIG. 12.14. Extensão do teorema de Stokes a superfícies que são imagens biunívocas de regiões multiplamente conexas.

11.12). Em vez do integral de linha que aparece em (12.27) necessitamos de uma soma de integrais de linha, com sinais adequados, tomados ao longo das imagens das curvas que constituem a fronteira de T . Por exemplo, se T tem dois buracos, como se mostra na figura 12.14, e se as curvas fronteira Γ , Γ_1 e Γ_2 são percorridos como se indica, a identidade no teorema de Stokes toma a forma

$$\iint_S (\text{rot } F) \cdot n \, dS = \oint_C F \cdot d\rho + \oint_{C_1} F \cdot d\rho_1 + \oint_{C_2} F \cdot d\rho_2,$$

com C , C_1 e C_2 as imagens de Γ , Γ_1 e Γ_2 , respectivamente, e ρ , ρ_1 e ρ_2 as funções compostas $\rho(t) = r[\gamma(t)]$, $\rho_1(t) = r[\gamma_1(t)]$, $\rho_2(t) = r[\gamma_2(t)]$. Aqui, γ , γ_1 e γ_2 são as funções que definem Γ , Γ_1 e Γ_2 nos sentidos indicados. As curvas C , C_1 e C_2 serão descritas nos sentidos induzidos pela aplicação r a partir de Γ , Γ_1 e Γ_2 .

O teorema de Stokes pode igualmente generalizar-se a algumas (mas não todas) superfícies regulares não simples. Vamos apresentar algumas das possibilidades através de alguns exemplos.

Consideremos em primeiro lugar o cilindro desenhado na figura 12.15. Pode considerar-se como a reunião de duas superfícies paramétricas regulares simples S_1 e S_2 , imagens de dois rectângulos adjacentes T_1 e T_2 , através das aplicações r_1 e r_2 , respectivamente. Se γ_1 descreve a fronteira Γ_1 , orientada positivamente, de T_1 e γ_2 a fronteira de Γ_2 igualmente orientada positivamente, as funções ρ_1 e ρ_2 definidas por

$$\rho_1(t) = r_1[\gamma_1(t)], \quad \rho_2(t) = r_2[\gamma_2(t)]$$

descrevem as imagens C_1 e C_2 de Γ_1 e Γ_2 , respectivamente. Neste exemplo as representações r_1 e r_2 podem escolher-se de modo que estejam de acordo na intersecção $\Gamma_1 \cap \Gamma_2$. Se aplicamos o teorema de Stokes a cada parte S_1 e S_2 e adicionamos membro a membro obtemos

$$(12.51) \quad \iint_{S_1} (\text{rot } F) \cdot n_1 \, dS + \iint_{S_2} (\text{rot } F) \cdot n_2 \, dS = \int_{C_1} F \cdot d\rho_1 + \int_{C_2} F \cdot d\rho_2,$$

com n_1 e n_2 as normais determinadas pelos produtos vectoriais fundamentais de r_1 e r_2 , respectivamente.

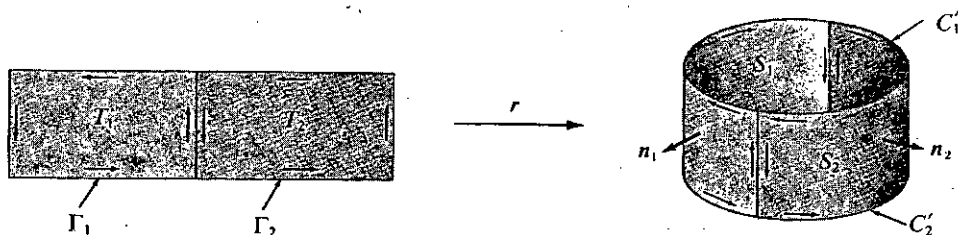


FIG. 12.15. Extensão do teorema de Stokes a um cilindro.

Designemos agora por r a aplicação de $T_1 \cup T_2$, que coincide com r_1 em T_1 e com r_2 em T_2 , e seja n a normal unitária correspondente determinada pelo produto vectorial fundamental de r . Uma vez que as normais n_1 e n_2 têm sentidos concordantes em $S_1 \cap S_2$, a normal unitária n é a mesma que n_1 em S_1 e que n_2 em S_2 . Consequentemente a soma dos integrais de superfície do primeiro membro de (12.51) é igual a

$$\iint_{S_1 \cup S_2} (\text{rot } F) \cdot n \, dS.$$

Neste exemplo, as representações r_1 e r_2 podem escolher-se de maneira que ρ_1 e ρ_2 definam sentidos *opostos* ao longo de cada arco da intersecção $C_1 \cap C_2$, como se indica na figura 12.15 através das setas. Os dois integrais de linha, no segundo membro de (12.51), podem substituir-se por uma soma de integrais de linha ao longo dos dois círculos C'_1 e C'_2 formando os bordos superior e inferior de $S_1 \cup S_2$, visto que os integrais de linha ao longo de cada arco da intersecção $C_1 \cap C_2$ anulam-se. Consequentemente, a equação (12.51) pode escrever-se

$$(12.52) \quad \iint_{S_1 \cup S_2} (\text{rot } F) \cdot n \, dS = \int_{C'_1} F \cdot d\rho_1 + \int_{C'_2} F \cdot d\rho_2,$$

onde, nos integrais de linha, C'_1 e C'_2 são descritas no sentido induzido por Γ_1 e Γ_2 . As duas circunferências C'_1 e C'_2 dizem-se formar a fronteira completa de $S_1 \cup S_2$. A equação (12.52) exprime o integral de superfície de $(\text{rot } F) \cdot n$ sobre $S_1 \cup S_2$ como um integral de linha sobre a fronteira completa de $S_1 \cup S_2$. Esta equação é a extensão do teorema de Stokes para um cilindro.

Suponhamos agora que aplicamos os mesmos conceitos à superfície representada na figura 12.16. Esta é ainda a reunião de duas superfícies paramétricas regulares simples S_1 e S_2 , as imagens de dois rectângulos adjacentes T_1 e T_2 . Esta superfície particular chama-se a *superfície de Möbius*[†]; pode construir-se facilmente um modelo de tal superfície recorrendo a uma tira rectangular de papel, dando a um dos lados extremos uma meia volta e justapondo em seguida esses lados extremos. Definem-se ρ_1 e ρ_2 , C_1 e C_2 para a banda de Möbius, tal como os definimos para o exemplo anterior do cilindro. O bordo de $S_1 \cup S_2$ neste caso é uma curva fechada simples C' , em vez de duas. Esta curva constitui a fronteira completa da superfície de Möbius.

Se aplicamos o teorema de Stokes a cada uma das partes S_1 e S_2 , como fizemos para o cilindro, obtemos (12.51). Mas se tentamos manter os dois integrais de superfície e os dois integrais de linha como atrás, encontramos duas dificuldades. A primeira, as duas normais n_1 e n_2 não têm sentidos concordantes em toda a intersecção $C_1 \cap C_2$. (Ver figura 12.16). Portanto não podemos definir uma normal n para a superfície, fazendo $n = n_1$ sobre S_1 e $n = n_2$ sobre S_2 , como fizemos para o cilindro. Isto não é,

[†] Devida a A. F. Möbius (1790-1868), um aluno de Gauss. Com 26 anos de idade foi nomeado professor de Astronomia em Leipzig, cargo que desempenhou até à sua morte. Trouxe algumas contribuições importantes à Mecânica Celeste, mas os seus trabalhos mais importantes dizem respeito à geometria e à teoria dos números.

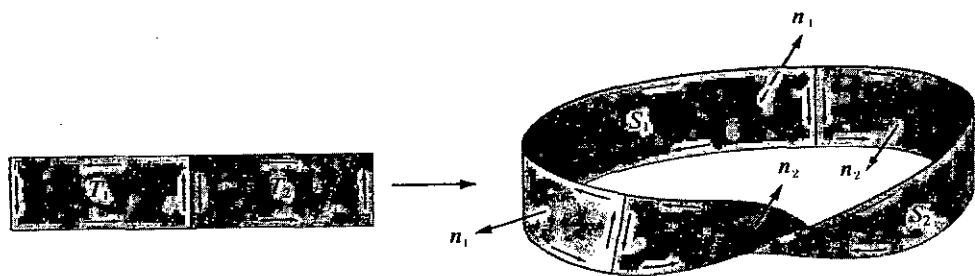


FIG. 12.16. A superfície de Möbius considerada como a reunião de duas superfícies paramétricas simples. O teorema de Stokes não se aplica a uma superfície de Möbius.

contudo, relevante, porque podemos definir \mathbf{n} como sendo \mathbf{n}_1 sobre S_1 e em $C_1 \cap C_2$, e depois definir \mathbf{n} como sendo \mathbf{n}_2 em todo o resto. Isto conduz-nos à definição de uma normal descontínua, mas as descontinuidades assim introduzidas formam um conjunto de medida nulo no plano OUV e não afectam a existência ou o valor do integral de superfície.

$$\iint_{S_1 \cup S_2} (\text{rot } \mathbf{F}) \cdot \mathbf{n} \, dS.$$

Uma dificuldade mais séria aparece quando tentamos manter os integrais de linha. Neste exemplo não é possível escolher as aplicações \mathbf{r}_1 e \mathbf{r}_2 de maneira que \mathbf{p}_1 e \mathbf{p}_2 definam sentidos opostos em cada uma das partes de $C_1 \cap C_2$. É o que se indica pelas setas na figura 12.16; uma dessas partes é descrita duas vezes no mesmo sentido. Sobre este arco os correspondentes integrais de linha não se anulam necessariamente um ao outro como aconteceu no caso do cilindro. Portanto a soma dos integrais de linha em (12.51) não é necessariamente igual ao integral de linha tomado ao longo da fronteira completa de $S_1 \cup S_2$, e o teorema de Stokes não pode estender-se à superfícies de Möbius.

Observação: O cilindro e a superfície de Möbius são exemplos de superfícies, respectivamente, *orientáveis* e *não orientáveis*. Não vamos definir estes termos aqui com rigor, mas vamos contudo mencionar algumas das suas diferenças. Para uma superfície orientável $S_1 \cup S_2$, formada por duas superfícies paramétricas regulares simples como as consideradas atrás, as aplicações \mathbf{r}_1 e \mathbf{r}_2 podem sempre ser escolhidas de maneira que \mathbf{p}_1 e \mathbf{p}_2 definam sentidos opostos em cada um dos arcos de intersecção $C_1 \cap C_2$. Para uma superfície não orientável tal escolha não é possível. Para uma superfície regular orientável pode definir-se um vector normal unitário de uma maneira contínua sobre toda a superfície. No caso de uma superfície não orientável tal definição não é possível. Um modelo em papel de uma superfície orientável tem sempre duas faces que podem diferenciar-se pintando-as com cores diferentes. As superfícies não orientáveis possuem unicamente uma face. Para um estudo rigoroso destas e de outras propriedades de superfícies orientáveis e não orientáveis, consul-

tar qualquer tratado de topologia combinatória. O teorema de Stokes pode generalizar-se a superfícies orientáveis por um processo análogo ao seguido para o caso do cilindro.

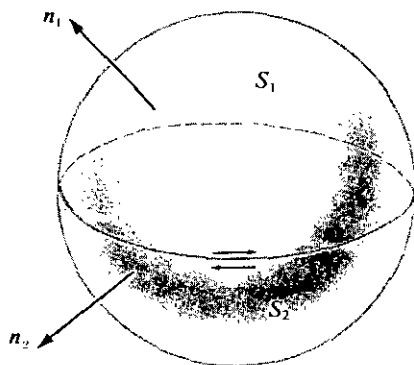


FIG. 12.17. Extensão do teorema de Stokes a uma esfera.

Outra superfície orientável é a esfera desenhada na figura 12.17. Esta superfície é a reunião de duas superfícies paramétricas simples (semi-esferas) S_1 e S_2 , as quais se podem considerar como imagens de um círculo do plano XOY através das aplicações r_1 e r_2 , respectivamente. Atribuímos a r , ρ_1 , ρ_2 , C_1 , C_2 o mesmo significado que nos exemplos anteriores. Neste caso as curvas C_1 e C_2 estão completamente identificadas pela aplicação r (coincidem ao longo do equador), e a superfície $S_1 \cup S_2$ diz-se *fechada*. Além disso, n_1 e n_2 podem escolher-se de maneira que os sentidos definidos por ρ_1 e ρ_2 sejam opostos em C_1 e C_2 , como se sugere pelas setas na figura 12.17. (O que é possível porque S_1 e S_2 são orientáveis). Se aplicarmos o teorema de Stokes a cada semi-esfera e adicionarmos as igualdades correspondentes, obtemos (12.51) como anteriormente. As normais n_1 e n_2 têm sentidos concordantes em $C_1 \cap C_2$ e podemos reunir os integrais sobre S_1 e S_2 num único integral sobre a esfera. Os dois integrais de linha do segundo membro de (12.51) anulam-se mutuamente, obtendo-se pois a fórmula

$$\iint_{S_1 \cup S_2} (\text{rot } F) \cdot n \, dS = 0.$$

Isto é válido não somente para a esfera, mas para qualquer superfície fechada orientável.

12.19 O teorema da divergência (teorema de Gauss)

O teorema de Stokes exprime uma relação entre um integral estendido a uma determinada superfície e o integral de linha tomado ao longo da curva ou curvas que constituem a fronteira da referida superfície. O teorema da divergência exprime uma relação

entre um integral triplo estendido a um sólido e um integral de superfície tomado sobre a superfície fronteira desse sólido.

TEOREMA 12.6. TEOREMA DA DIVERGÊNCIA. *Sejam V um sólido no espaço tridimensional limitado por uma superfície fechada orientável S e \mathbf{n} a normal unitária para o exterior de S . Se \mathbf{F} é um campo vectorial continuamente diferenciável definido em V , tem-se*

$$(12.53) \quad \iiint_V (\operatorname{div} \mathbf{F}) \, dx \, dy \, dz = \iint_S \mathbf{F} \cdot \mathbf{n} \, dS.$$

Nota: Exprimindo \mathbf{F} e \mathbf{n} em função das respectivas componentes, por exemplo

$$\mathbf{F}(x, y, z) = P(x, y, z)\mathbf{i} + Q(x, y, z)\mathbf{j} + R(x, y, z)\mathbf{k}$$

e

$$\mathbf{n} = \cos \alpha \mathbf{i} + \cos \beta \mathbf{j} + \cos \gamma \mathbf{k},$$

(12.53) pode escrever-se

$$(12.54) \quad \iiint_V \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z} \right) dx \, dy \, dz = \iint_S (P \cos \alpha + Q \cos \beta + R \cos \gamma) \, dS.$$

Demonstração. Basta estabelecer as três seguintes igualdades

$$\iiint_V \frac{\partial P}{\partial x} \, dx \, dy \, dz = \iint_S P \cos \alpha \, dS,$$

$$\iiint_V \frac{\partial Q}{\partial y} \, dx \, dy \, dz = \iint_S Q \cos \beta \, dS,$$

$$\iiint_V \frac{\partial R}{\partial z} \, dx \, dy \, dz = \iint_S R \cos \gamma \, dS,$$

e somá-los membro a membro para se obter (12.54). Começemos pela terceira e vamos demonstrá-la para sólidos de um tipo muito especial.

Admitamos que V é um conjunto de pontos (x, y, z) satisfazendo a uma relação da forma

$$g(x, y) \leq z \leq f(x, y) \quad \text{para } (x, y) \text{ em } T,$$

sendo T uma região conexa de XOY e f e g funções contínuas em T com $g(x, y) \leq f(x, y)$ para todo (x, y) em T . Geometricamente, isto significa que T é a projecção de V no plano XOY . Toda a recta paralela ao eixo OZ que atravessa T intersecta o sólido V ao longo de um segmento de recta que une a superfície $z = g(x, y)$ com a superfície $z = f(x, y)$. A superfície fronteira S é formada pela parte superior S_1 , definida pela fórmula explícita $z = f(x, y)$; a parte inferior S_2 , definida por $z = g(x, y)$; e (possivel-

mente) por uma parte S_3 de uma superfície cilíndrica gerada por uma paralela ao eixo OZ movendo-se ao longo da fronteira de T . A normal exterior a S tem componente segundo OZ não negativa em S_1 , não positiva em S_2 e é paralela a XOY em S_3 . Sólidos deste tipo dizem-se "projectáveis em XOY ". (Na figura 12.18 apresenta-se um exemplo). Incluem-se neste tipo todos os sólidos convexos (por exemplo esferas, elipsoides, cubos) e muitos outros sólidos que não são convexos (como por exemplo o tipo de eixo paralelo a OZ).

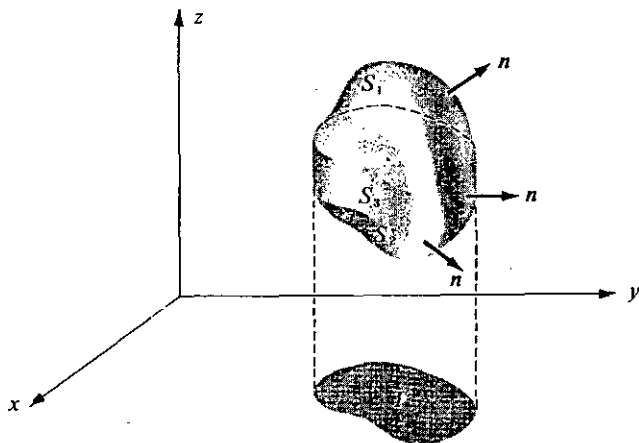


FIG. 12.18. Exemplo de sólido projectáveis $-xy$.

A ideia para a demonstração é muito simples. Expressimos o integral triplo como um integral duplo estendido à projecção T . Mostramos em seguida que este integral duplo tem o mesmo valor que o integral de superfície considerado no enunciado.

Consideremos então a fórmula

$$\iiint_V \frac{\partial R}{\partial z} dx dy dz = \iint_T \left[\int_{g(x,y)}^{f(x,y)} \frac{\partial R}{\partial z} dz \right] dx dy.$$

O integral a respeito de z pode calcular-se mediante o segundo teorema fundamental do cálculo, dando-nos

$$(12.55) \quad \iiint_V \frac{\partial R}{\partial z} dx dy dz = \iint_T \{R[x, y, f(x, y)] - R[x, y, g(x, y)]\} dx dy.$$

Para o integral de superfície podemos escrever

$$(12.56) \quad \iint_S R \cos \gamma dS = \iint_{S_1} R \cos \gamma dS + \iint_{S_2} R \cos \gamma dS + \iint_{S_3} R \cos \gamma dS.$$

Em S_3 a normal \mathbf{n} é paralela ao plano XOY , pelo que $\cos \gamma = 0$ e o integral estendido a S_3 é nulo. Sobre a superfície S_1 recorremos à representação

$$\mathbf{r}(x, y) = xi + yj + f(x, y)k,$$

e sobre S_2

$$\mathbf{r}(x, y) = xi + yj + g(x, y)k.$$

Sobre S_1 a normal \mathbf{n} tem o mesmo sentido que o produto vectorial $\partial \mathbf{r} / \partial x \times \partial \mathbf{r} / \partial y$, pelo que podemos escrever [ver equação (12.25)]

$$\iint_{S_1} R \cos \gamma \, dS = \iint_{S_1} R \, dx \wedge dy = \iint_T R[x, y, f(x, y)] \, dx \, dy.$$

Sobre S_2 a normal \mathbf{n} tem o sentido oposto de $\partial \mathbf{r} / \partial x \times \partial \mathbf{r} / \partial y$ pelo que, devido a (12.26), se terá

$$\iint_{S_2} R \cos \gamma \, dS = - \iint_{S_2} R \, dx \wedge dy = - \iint_T R[x, y, g(x, y)] \, dx \, dy.$$

Deste modo (12.56) vem

$$\iint_S R \cos \gamma \, dS = \iint_T \{R[x, y, f(x, y)] - R[x, y, g(x, y)]\} \, dx \, dy.$$

Comparando-a com (12.55) vemos que

$$\iiint_V \frac{\partial R}{\partial z} \, dx \, dy \, dz = \iint_S R \cos \gamma \, dS.$$

Na demonstração precedente a hipótese de que V é projectável em XOY permite-nos exprimir o integral triplo sobre V como um integral duplo sobre a sua projecção T no plano XOY . É evidente que se V for "projectável em YOZ " podemos utilizar o mesmo tipo de argumento para demonstrar que

$$\iiint_V \frac{\partial P}{\partial x} \, dx \, dy \, dz = \iint_S P \cos \alpha \, dS;$$

e se V for "projectável em XOZ " obtemos

$$\iiint_V \frac{\partial Q}{\partial y} \, dx \, dy \, dz = \iint_S Q \cos \beta \, dS.$$

Concluimos assim que o teorema da divergência é verdadeiro para todos os sólidos projectáveis sobre os três planos coordenados. Em particular o teorema é válido para todo o sólido conexo.

Um toro de eixo paralelo a OZ é "projectável em XOY ", mas não em XOZ ou YOZ . Para generalizar o teorema da divergência a um tal sólido dividimos o toro em quatro partes iguais por planos que passam pelo seu eixo e são paralelos aos planos XOZ e YOZ , respectivamente, e aplicamos o teorema da divergência a cada parte. O integral triplo estendido a todo o toro é a soma dos integrais triplos estendidos a cada uma das quatro partes. Quando adicionamos os integrais de superfície estendidos às quatro partes encontramos que as contribuições das faces comuns às partes adjacentes se anulam mutuamente, uma vez que as normais exteriores tem sentidos opostos sobre cada par de tais faces. Deste modo a soma dos integrais de superfície relativas às quatro partes é igual ao integral de superfície estendido a todo o toro. Este exemplo ilustra o modo como o teorema de divergência pode estender-se a certos sólidos não convexos.

12.20 Aplicações do teorema da divergência

Os conceitos do rotacional e divergência de um campo vectorial $F = Pi + Qj + Rk$ foram introduzidos na Secção 12.12 pelas fórmulas

$$(12.57) \quad \operatorname{div} F = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}$$

e

$$(12.58) \quad \operatorname{rot} F = \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z} \right) i + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x} \right) j + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) k.$$

Para se calcular $\operatorname{div} F$ e $\operatorname{rot} F$ a partir destas fórmulas necessita-se o conhecimento das componentes de F . Estas componentes, por sua vez, dependem da escolha dos eixos coordenados no espaço tridimensional. Uma mudança de eixos significará uma mudança nas componentes de F e, presumivelmente, uma correspondente mudança nas funções $\operatorname{div} F$ e $\operatorname{rot} F$. Com o auxílio do teorema de Stokes e do teorema da divergência podemos obter fórmulas para a divergência e rotacional nos quais não intervêm as componentes de F . Estas fórmulas mostram que o rotacional e a divergência constituem propriedades intrínsecas do campo vectorial F e não dependem da escolha particular dos eixos coordenados. Analisemos em primeiro lugar a fórmula para a divergência.

TEOREMA 12.7. *Seja $V(t)$ uma esfera de raio $t > 0$ com centro no ponto a do espaço tridimensional, e represente $S(t)$ a fronteira de $V(t)$. Seja F um campo vectorial continuamente diferenciável em $V(t)$. Então se $|V(t)|$ representa o volume de $V(t)$, e se n é a normal unitária exterior a S , tem-se*

$$(12.59) \quad \operatorname{div} F(a) = \lim_{t \rightarrow 0} \frac{1}{|V(t)|} \iint_{S(t)} F \cdot n \, dS.$$

Demonstração. Seja $\varphi = \operatorname{div} F$. Dado $\epsilon > 0$ temos que determinar um $\delta > 0$ tal que

$$(12.60) \quad \left| \varphi(a) - \frac{1}{|V(t)|} \iint_{S(t)} \mathbf{F} \cdot \mathbf{n} \, dS \right| < \epsilon \quad \text{sempre que } 0 < t < \delta.$$

Visto que φ é contínua em a , para o ϵ dado existe uma 3-bola $B(a; h)$ tal que

$$|\varphi(x) - \varphi(a)| < \frac{\epsilon}{2} \quad \text{sempre que } x \in B(a; h).$$

Portanto, se escrevermos $\varphi(a) = \varphi(x) + [\varphi(a) - \varphi(x)]$ e integramos ambos os membros desta igualdade sobre a esfera $V(t)$ de raio $t < h$, encontramos

$$\varphi(a) |V(t)| = \iiint_{V(t)} \varphi(x) \, dx \, dy \, dz + \iiint_{V(t)} [\varphi(a) - \varphi(x)] \, dx \, dy \, dz.$$

Se aplicamos o teorema de divergência ao primeiro integral triplo do segundo membro e transpomos esse termo para o primeiro membro, obtemos a relação

$$\left| \varphi(a) |V(t)| - \iint_{S(t)} \mathbf{F} \cdot \mathbf{n} \, dS \right| \leq \iiint_{V(t)} |\varphi(a) - \varphi(x)| \, dx \, dy \, dz \leq \frac{\epsilon}{2} |V(t)| < \epsilon |V(t)|.$$

Quando dividimos esta desigualdade por $|V(t)|$ vemos que (12.60) é válida com $\delta = h$, o que demonstra o teorema.

Na demonstração anterior não fizemos qualquer uso especial do facto de que $V(t)$ era uma esfera. O mesmo teorema permanece verdadeiro se, em vez de esferas, utilizarmos qualquer conjunto de sólidos $V(t)$ para os quais o teorema da divergência seja válido, com tanto que estes sólidos contenham o ponto a e tendam para a quando $t \rightarrow 0$. Por exemplo, cada $V(t)$ pode ser um cubo inscrito numa esfera de raio t em torno de a ; aplicar-se-ia exactamente a mesma definição.

O teorema 12.7 pode ser utilizado para dar uma interpretação física da divergência. Suponhamos que \mathbf{F} representa o vector densidade de fluxo de uma corrente estacionária. Então o integral de superfície $\iint_{S(t)} \mathbf{F} \cdot \mathbf{n} \, dS$ mede a massa total do fluido passando através de S na unidade de tempo e no sentido de \mathbf{n} . O cociente $\iint_{S(t)} \mathbf{F} \cdot \mathbf{n} \, dS / |V(t)|$ representa a massa por unidade de volume que passa através de S na unidade de tempo e no sentido de \mathbf{n} . Quanto $t \rightarrow 0$, o limite deste cociente é a divergência de \mathbf{F} em a . Por isso, a divergência em a pode ser interpretada como o coeficiente de variação da massa, por unidade de volume, por unidade de tempo, em a .

Em alguns livros de análise vectorial, a equação (12.59) é tomada como *definição de divergência*. Tal facto permite imediatamente atribuir um significado físico à divergência. Além disso, na fórmula (12.59) não intervêm as componentes de \mathbf{F} . Portanto ela é válida em qualquer sistema de coordenadas. Se escolhemos para $V(t)$ um cubo com as arestas paralelas aos eixos coordenados e o centro em a , podemos servir-nos de (12.59) para deduzir (12.57) que exprime $\text{div } \mathbf{F}$ e função das componentes de \mathbf{F} . Este procedimento esboça-se no Exercício 14 da Secção 12.21.

Existe uma fórmula análoga a (12.59) que é muitas vezes usada como uma alternativa na definição de rotacional, e que se escreve

$$(12.61) \quad \text{rot } \mathbf{F}(\mathbf{a}) = \lim_{t \rightarrow 0} \frac{1}{|V(t)|} \iint_{S(t)} \mathbf{n} \times \mathbf{F} dS,$$

onde $V(t)$ e $S(t)$ têm o mesmo significado que no teorema 12.7. O integral de superfície que figura no segundo membro tem um integrando vectorial. Tais integrais podem definir-se por intermédio das componentes. A demonstração de (12.61) é análoga à do teorema 12.7.

Existe outra fórmula em que intervem o rotacional, a qual pode ser deduzida de (12.61) ou ser estabelecida independentemente, a qual estabelece que

$$(12.62) \quad \mathbf{n} \cdot \text{rot } \mathbf{F}(\mathbf{a}) = \lim_{t \rightarrow 0} \frac{1}{|S(t)|} \oint_{C(t)} \mathbf{F} \cdot d\boldsymbol{\alpha}.$$

Nesta fórmula, $S(t)$ é um círculo de raio t e centro \mathbf{a} , e $|S(t)|$ representa a sua área. O vector \mathbf{n} é a normal unitária a $S(t)$, e $\boldsymbol{\alpha}$ a função que define $C(t)$ num sentido tal quando observado de \mathbf{n} parece positivo. O campo vectorial \mathbf{F} é suposto continuamente diferenciável em $S(t)$. Uma demonstração de (12.62) pode ser estabelecida pelo mesmo método usado na demonstração de (12.59). Fazemos $\varphi(\mathbf{x}) = \mathbf{n} \cdot \text{rot } \mathbf{F}(\mathbf{x})$ e raciocinamos como atrás, excepto em que usamos integrais de superfície em vez de integrais triplos e o teorema de Stokes em vez do teorema da divergência.

Se \mathbf{F} é um campo de velocidade, o integral de linha ao longo de $C(t)$ chama-se a circulação de \mathbf{F} ao longo de $C(t)$; o limite em (12.62) representa a circulação por unidade de área no ponto \mathbf{a} . Assim, $\mathbf{n} \cdot \text{rot } \mathbf{F}(\mathbf{a})$ pode ser considerada como uma "densidade de circulação" de \mathbf{F} em \mathbf{a} , com respeito ao plano perpendicular a \mathbf{n} em \mathbf{a} .

Quando \mathbf{n} toma sucessivamente os valores \mathbf{i} , \mathbf{j} , \mathbf{k} , os produtos escalares $\mathbf{i} \cdot \text{rot } \mathbf{F}$, $\mathbf{j} \cdot \text{rot } \mathbf{F}$, e $\mathbf{k} \cdot \text{rot } \mathbf{F}$ são as componentes do $\text{rot } \mathbf{F}$ no sistema cartesiano rectangular. Quando a equação (12.61) é tomada como ponto de partida para a definição de rotacional, a fórmula (12.58) para as componentes rectangulares de $\text{rot } \mathbf{F}$ pode ser deduzida de (12. 2) exactamente da mesma maneira.

12.21 Exercícios

1. Seja S a superfície do cubo, $0 \leq x \leq 1$, $0 \leq y \leq 1$, $0 \leq z \leq 1$, e \mathbf{n} a normal unitária dirigida para o exterior de S . Se $\mathbf{F}(x, y, z) = x^2\mathbf{i} + y^2\mathbf{j} + z^2\mathbf{k}$, aplicar o teorema da divergência para calcular o integral de superfície $\iint_S \mathbf{F} \cdot \mathbf{n} dS$. Verificar o resultado através do cálculo directo

do integral de superfície dado.

2. A esfera $x^2 + y^2 + z^2 = 25$ é intersectada pelo plano $z = 3$. A parte menor constitui um sólido V limitado por uma superfície fechada S_0 formada de duas partes, uma parte esférica S_1 e uma parte plana S_2 . Se a normal unitária exterior de V é $\cos \alpha \mathbf{i} + \cos \beta \mathbf{j} + \cos \gamma \mathbf{k}$, calcular o valor do integral de superfície

$$\iint_S (xz \cos \alpha + yz \cos \beta + \cos \gamma) dS$$

se (a) S é a parte esférica S_1 , (b) S é a base plana S_2 , (c) S é toda a superfície fronteira S_0 . Resolver a alínea (c) com os resultados das alíneas (a) e (b), e também por aplicação da teorema da divergência.

3. Seja $\mathbf{n} = \cos \alpha \mathbf{i} + \cos \beta \mathbf{j} + \cos \gamma \mathbf{k}$ a normal unitária exterior à superfície fechada S que limita um sólido homogêneo V do tipo considerado no teorema da divergência. Admita-se que se conhecem o centro de massa $(\bar{x}, \bar{y}, \bar{z})$ e o volume $|V|$ de V . Calcular os seguintes integrais de superfície em função de $|V|$ e $\bar{x}, \bar{y}, \bar{z}$.

(a) $\iint_S (x \cos \alpha + y \cos \beta + z \cos \gamma) dS$.

(b) $\iint_S (xz \cos \alpha + 2yz \cos \beta + 3z^2 \cos \gamma) dS$.

(c) $\iint_S (y^2 \cos \alpha + 2xy \cos \beta - xz \cos \gamma) dS$.

(d) Expressar $\iint_S (x^2 + y^2) (\mathbf{x}\mathbf{i} + \mathbf{y}\mathbf{j}) \cdot \mathbf{n} dS$ em função do volume $|V|$ e de um momento de inércia do sólido.

Nos Exercícios 4 a 10, $\partial f / \partial n$ e $\partial g / \partial n$ representam derivadas direccionais dos campos escalares f e g na direcção da normal unitária \mathbf{n} , exterior a uma superfície fechada S que limita um sólido V do tipo considerado no teorema da divergência. Visto que $\partial f / \partial n = \nabla f \cdot \mathbf{n}$ e $\partial g / \partial n = \nabla g \cdot \mathbf{n}$. Em cada um dos exercícios provar a proposição enunciada. Deve admitir-se a continuidade de todas as derivadas que se considerem.

4. $\iint_S \frac{\partial f}{\partial n} dS = \iiint_V \nabla^2 f dx dy dz$.

5. $\iint_S \frac{\partial f}{\partial n} dS = 0$ sempre que f seja harmónica em V .

6. $\iint_S f \frac{\partial g}{\partial n} dS = \iiint_V f \nabla^2 g dx dy dz + \iiint_V \nabla f \cdot \nabla g dx dy dz$.

7. $\iint_S \left(f \frac{\partial g}{\partial n} - g \frac{\partial f}{\partial n} \right) dS = \iiint_V (f \nabla^2 g - g \nabla^2 f) dx dy dz$.

8. $\iint_S f \frac{\partial g}{\partial n} dS = \iint_S g \frac{\partial f}{\partial n} dS$ se f e g são ambas harmónicas em V .

9. $\iint_S f \frac{\partial f}{\partial n} dS = \iiint_V |\nabla f|^2 dx dy dz$ se f é harmónica em V .

10. $\nabla^2 f(\mathbf{a}) = \lim_{t \rightarrow 0} \frac{1}{|V(t)|} \iint_{S(t)} \frac{\partial f}{\partial n} dS$, onde $V(t)$ é o sólido esférico de raio t e centro \mathbf{a} , $S(t)$ é a superfície de $V(t)$, e $|V(t)|$ é o volume de $V(t)$.
11. Seja V uma região convexa do espaço tridimensional cuja fronteira é uma superfície fechada S e seja \mathbf{n} a normal unitária exterior de S . Representem \mathbf{F} e \mathbf{G} dois campos vectoriais continuamente diferenciáveis tais que

$$\operatorname{rot} \mathbf{F} = \operatorname{rot} \mathbf{G} \quad \text{e} \quad \operatorname{div} \mathbf{F} = \operatorname{div} \mathbf{G} \quad \text{em todo } V,$$

e ainda

$$\mathbf{G} \cdot \mathbf{n} = \mathbf{F} \cdot \mathbf{n} \quad \text{em toda a superfície } S.$$

Demonstrar que $\mathbf{F} = \mathbf{G}$ em todo V . [Sugestão: Sendo $\mathbf{H} = \mathbf{F} - \mathbf{G}$, determinar um campo escalar f tal que $\mathbf{H} = \Delta f$, e utilizar uma identidade conveniente para provar que $\iiint_V \|\nabla f\|^2 dx dy dz = 0$. A partir daqui demonstrar que $\mathbf{H} = \mathbf{0}$ em V .]

12. Dado um campo vectorial \mathbf{G} e dois campos escalares f e g , cada um deles continuamente diferenciável num sólido convexo V limitado por uma superfície fechada S . Seja \mathbf{n} a normal unitária exterior a S . Provar que existe quando muito um campo vectorial \mathbf{F} verificando as três condições seguintes:

$$\operatorname{rot} \mathbf{F} = \mathbf{G} \quad \text{e} \quad \operatorname{div} \mathbf{F} = g \quad \text{em } V, \quad \mathbf{F} \cdot \mathbf{n} = f \quad \text{sobre } S.$$

13. Seja S uma superfície paramétrica regular gozando da propriedade de que cada recta passando por um ponto P intersecta S quando muito uma vez. Seja $\Omega(S)$ o conjunto das rectas passando por P e intersectando S . (Ver figura 12.19). O conjunto $\Omega(S)$ chama-se *ângulo sólido* de vértice P subtendido por S . Seja $\Sigma(a)$ a intersecção de $\Omega(S)$ com a superfície da esfera de raio a e centro P . O cociente

$$\frac{\text{área de } \Sigma(a)}{a^2}$$

representa-se por $|\Omega(S)|$ e usa-se como uma medida do ângulo sólido $\Omega(S)$.

FIG. 12.19. O ângulo sólido $\Omega(S)$ com vértice em P subtende a uma superfície S .

A sua medida é definida pelo cociente $|\Omega(S)| = \frac{\text{área de } \Sigma(a)}{a^2}$.

- (a) Demonstrar que este cociente é igual ao integral de superfície

$$\iint_S \frac{\mathbf{r} \cdot \mathbf{n}}{r^3} dS,$$

onde \mathbf{r} é o raio vector de P para um ponto arbitrário de S , e $r = \|\mathbf{r}\|$. O vector \mathbf{n} é a normal unitária a S dirigida para o lado contrário ao de P . Isto mostra que o cociente $|\Omega(S)|$ é independente do raio a . Consequentemente o ângulo sólido pode medir-se pela área da intersecção de $\Omega(S)$ com a esfera de raio unidade e centro em P . [Sugestão: Aplicar o teorema da divergência à parte de $\Omega(S)$ compreendida entre S e $\Sigma(a)$.]

(b) Dois planos intersectam-se ao longo do diâmetro de uma esfera com centro em P . O ângulo de intersecção é θ , com $0 < \theta < \pi$. Seja S a parte menor da superfície da esfera intersectada pelos dois planos. Provar que $|\Omega(S)| = 2\theta$.

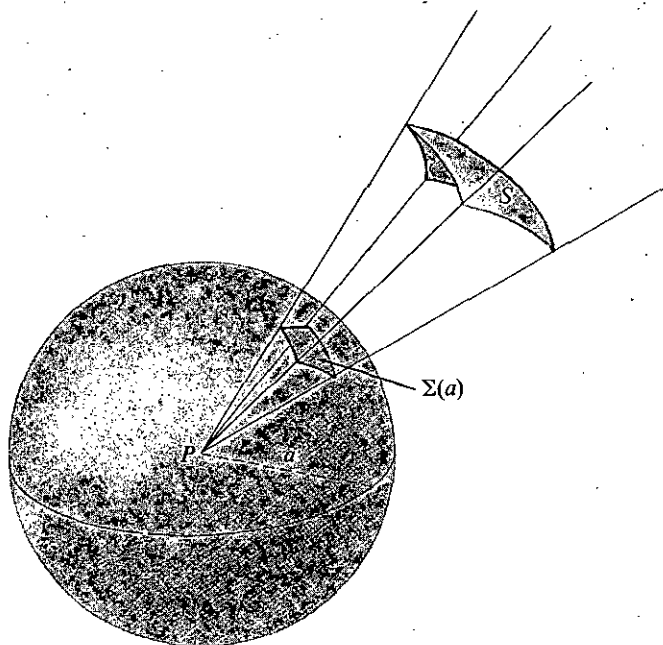


FIG. 12.19. O ângulo sólido $\Omega(S)$ com vértice em P subtensa a uma superfície S .

A sua medida é definida pelo cociente $|\Omega(S)| = \frac{\text{área de } \Sigma(a)}{a^2}$

14. Seja $V(t)$ um cubo de aresta $2t$ e centro \mathbf{a} , e represente $S(t)$ a sua superfície fronteira. Designe-se por \mathbf{n} a normal unitária exterior a $S(t)$ e seja $|V(t)|$ o volume do cubo. Para um dado campo vectorial \mathbf{F} que é continuamente diferenciável em \mathbf{a} , admita-se que existe o limite

$$\lim_{t \rightarrow 0} \frac{1}{|V(t)|} \iint_{S(t)} \mathbf{F} \cdot \mathbf{n} \, dS,$$

e utilize-se este limite como a definição da divergência $\text{div } \mathbf{F}(\mathbf{a})$. Escolham-se os eixos coordenados $OXYZ$ paralelos às arestas de $V(t)$ e sejam P , Q e R as componentes de \mathbf{F} nesse sistema de coordenadas. Provar que $\text{div } \mathbf{F}(\mathbf{a}) = D_1 P(\mathbf{a}) + D_2 Q(\mathbf{a}) + D_3 R(\mathbf{a})$. [Sugestão: Expressar o integral de superfície como uma soma de seis integrais duplos tomados sobre as faces do cubo. Provar depois que $1/|V(t)|$ vezes a soma dos dois integrais duplos estendidos às faces perpendiculares ao eixo OZ tende para o limite $D_3 R(\mathbf{a})$ quando $t \rightarrow 0$. Raciocinar de maneira análoga para os restantes termos].

15. Um campo escalar φ que é não nulo goza das propriedades

$$\|\nabla \varphi\|^2 = 4\varphi \quad \text{e} \quad \text{div}(\varphi \nabla \varphi) = 10\varphi.$$

Calcular o integral de superfície

$$\iint_S \frac{\partial \varphi}{\partial n} dS,$$

onde S é a superfície de uma esfera unitária com centro na origem, e $\partial \varphi / \partial n$ é a derivada direccional de φ na direcção da normal unitária exterior a S .

PARTE 3

TÓPICOS ESPECIAIS

FUNÇÕES DE CONJUNTO E PROBABILIDADE ELEMENTAR

13.1. Introdução histórica

Uma disputa entre jogadores em 1654 levou à criação, por dois famosos matemáticos franceses, Blaise Pascal e Pierre de Fermat, de uma teoria matemática das probabilidades. Antoine Gombard, Cavaleiro de Méré, um nobre francês interessado em questões de jogos e apostas, chamou a atenção de Pascal para uma aparente contradição num popular “jogo de dados”. O jogo consistia em lançar um par de dados 24 vezes; o problema estava em decidir se era correcto apostar a mesma importância a favor ou contra o aparecimento de pelo menos um “duplo seis”, nos vinte e quatro lançamentos. Uma regra de jogo, aparentemente bem estabelecida, conduziu de Méré a acreditar que seria vantajoso apostar por um “duplo seis” nos 24 lançamentos, mas os seus próprios cálculos mostravam-lhe o contrário.

Este e outros problemas postos por de Méré motivaram uma troca de correspondência entre Pascal e Fermat, na qual se estabeleceram pela primeira vez os princípios fundamentais da teoria das probabilidades. Embora alguns problemas especiais relativos a jogos de azar tivessem sido resolvidos por alguns matemáticos italianos nos séculos XV e XVI, não tinha sido desenvolvida qualquer teoria geral antes dessa famosa correspondência. O cientista Christian Huygens, um professor de Leibniz, conhecedor do conteúdo dessa correspondência publicou (em 1657) o primeiro livro sobre a teoria das probabilidades; intitulado *De Ratiociniis in Ludo Aleae*, era um tratado com problemas associados com os jogos. A teoria de probabilidades tornou-se rapidamente popular, devido às suas alusões aos “jogos de azar” e desenvolveu-se rapidamente durante o século XVIII. Quem mais contribuiu para o seu desenvolvimento, durante esse período, foram Jacob Bernoulli† (1654-1705) e Abraham de Moivre (1667-1754).

† Algumas vezes citado como James Bernoulli.

Em 1812 Pierre de Laplace (1749-1827) introduziu um grande número de ideias e técnicas matemáticas no seu livro *Théorie Analytique des Probabilités*. Antes de Laplace, a teoria das probabilidades consistia praticamente na análise matemática dos jogos de azar. Laplace provou que essa teoria das probabilidades tinha aplicações em muitos problemas científicos e práticos. *A teoria dos erros, a análise actuarial, e a mecânica estatística* são exemplos de algumas das aplicações importantes da teoria das probabilidades, desenvolvidas no século XIX.

Analogamente ao que se passou com muitos outros ramos da matemática, o desenvolvimento da teoria das probabilidades foi estimulado pela variedade das suas aplicações. Inversamente, cada avanço na teoria alargava o âmbito da sua influência. A estatística matemática é um dos exemplos mais importantes de aplicação da teoria das probabilidades; outras aplicações ocorrem em campos tão diferenciados como a genética, a psicologia, a economia e a engenharia. São muitos os autores que contribuíram com os seus resultados para a teoria das probabilidades desde o tempo de Laplace; entre os mais importantes contam-se Chebyshev, Markov, von Mises e Kolmogorov.

Uma das dificuldades no desenvolvimento de uma teoria matemática das probabilidades foi a de conseguir uma definição de probabilidades suficientemente rigorosa para a sua utilização matemática, mas suficientemente ampla para que seja aplicável a uma classe de fenómenos também, o mais amplexível. A pesquisa de uma definição completamente aceitável durou cerca de três séculos e foi caracterizada por muita controvérsia. O assunto foi definitivamente resolvido no século XX pelo tratamento da teoria das probabilidades de uma maneira axiomática. Em 1933, uma monografia do matemático russo A. Kolmogorov esboçava uma aproximação axiomática que constitui a base para a moderna teoria. (A monografia de Kolmogorov encontra-se traduzida em inglês, *Foundations of Probability Theory*, Chelsea, N.Y., 1950). Desde então as ideias refinaram-se mais e a teoria das probabilidades é agora parte de uma teoria mais geral conhecida por teoria da medida.

Neste capítulo apresentamos as noções básicas de uma teoria moderna das probabilidades, conjuntamente com as suas ligações com a teoria da medida. Dão-se também algumas aplicações, fundamentalmente à teoria dos jogos de azar, tais como lançamento de moedas, dados e jogos de cartas. Esta introdução pretende pôr em evidência a estrutura lógica da matéria como ciência dedutiva e motivar o interesse do leitor pelo pensamento probabilístico.

13.2. Funções de conjunto completamente aditivas

A área de uma região, o comprimento de uma curva, ou a massa de um sistema de partículas é um número que mede o tamanho ou o conteúdo de um conjunto. Todas estas medidas têm certas propriedades em comum. Estabelecidas de forma abstracta, conduzem a um conceito mais geral o de *função de conjunto finitamente aditiva*. Mais adiante introduziremos a probabilidade como outro exemplo de uma função deste tipo. Para preparar o caminho, analisamos primeiramente algumas propriedades comuns a todas estas funções.

Uma função $f: \mathcal{A} \rightarrow \mathbf{R}$, cujo domínio é uma família \mathcal{A} de conjuntos e cujos valores da função são números reais, diz-se uma *função de conjunto*. Se A é um conjunto da classe \mathcal{A} o valor da função em A representa-se por $f(A)$.

DEFINIÇÃO DE UMA FUNÇÃO DE CONJUNTO COMPLETAMENTE ADITIVA. Uma função de conjunto $f: \mathcal{A} \rightarrow \mathbf{R}$ diz-se *completamente aditiva* se

$$(13.1) \quad f(A \cup B) = f(A) + f(B)$$

sempre que A e B sejam conjuntos disjuntos em \mathcal{A} , tais que $A \cup B$ pertença também a \mathcal{A} .

A área, o comprimento e a massa são todos exemplos de funções de conjunto completamente aditivas. Nesta secção analisam-se consequências de (13.1).

Nas aplicações usuais, os conjuntos de \mathcal{A} são subconjuntos de um dado conjunto S , chamado o conjunto universal. É frequente ter que efectuar as operações de união, intersecção e complementação sobre os conjuntos de \mathcal{A} . Para se assegurar que \mathcal{A} é fechada a respeito destas operações restringimos \mathcal{A} a ser uma *Álgebra de Boole*, que se define do modo seguinte:

DEFINIÇÃO DE UMA ÁLGEBRA DE BOOLE DE CONJUNTOS. Uma classe não vazia \mathcal{A} de subconjuntos de um dado conjunto universal S diz-se uma *álgebra de Boole* se para todo A e B em \mathcal{A} tem

$$A \cup B \in \mathcal{A} \quad \text{e} \quad A' \in \mathcal{A}.$$

com $A' = S - A$ o complemento de A relativamente a S .

Uma álgebra de Boole \mathcal{A} é também fechada a respeito das intersecções e diferenças, uma vez que se tem

$$A \cap B = (A' \cup B')' \quad \text{e} \quad A - B = A \cap B'.$$

Isto implica que o conjunto vazio \emptyset pertence a \mathcal{A} , visto que $\emptyset = A - A$ para algum A em \mathcal{A} . Também o conjunto universal S pertence a \mathcal{A} visto que $S = \emptyset'$.

A partir dos subconjuntos de um conjunto universal dado S podem construir-se grande número de álgebras de Boole. A menor dessas álgebras é a classe $\mathcal{A}_0 = \{\emptyset, S\}$, a qual é formada unicamente pelos dois conjuntos especiais: \emptyset e S . No outro extremo está a classe \mathcal{A}_1 , que é formada por todos os subconjuntos de S . Toda a álgebra de Boole \mathcal{A} constituída por subconjuntos de S satisfaz às relações de inclusão $\mathcal{A}_0 \subseteq \mathcal{A} \subseteq \mathcal{A}_1$.

A propriedade da aditividade finita de funções de conjunto na Equação (13.1) exige que A e B sejam disjuntos. Destas exigências resulta o teorema seguinte:

TEOREMA 13.1. *Se $f: \mathcal{A} \rightarrow \mathbf{R}$ é uma função de conjunto finitamente aditiva definida sobre uma álgebra de Boole \mathcal{A} de conjuntos, então para todo o par de conjuntos A e B de \mathcal{A} tem-se*

$$(13.2) \quad f(A \cup B) = f(A) + f(B - A),$$

$$(13.3) \quad f(A \cup B) = f(A) + f(B) - f(A \cap B).$$

Demonstração. Os conjuntos A e $B - A$ são disjuntos e a sua união é $A \cup B$. Logo, aplicando (13.1) a A e $B - A$ obtemos (13.2).

Para demonstrar (13.3) observe-se que $A \cap B'$ e B são conjuntos disjuntos cuja união é $A \cup B$. Consequentemente, por (13.1) temos

$$(13.4) \quad f(A \cup B) = f(A \cap B') + f(B).$$

Também $A \cap B'$ e $A \cap B$ são conjuntos disjuntos cuja união é A , pelo que (13.1) nos dá

$$(13.5) \quad f(A) = f(A \cap B') + f(A \cap B).$$

Subtraindo (13.5) de (13.4) obtemos (13.3).

13.3. Medidas finitamente aditivas

As funções de conjunto que representam áreas, comprimentos, e massas têm várias propriedades em comum. Por exemplo, são todas funções de conjunto *não negativas*. Quer dizer,

$$f(A) \geq 0$$

para todo A da classe \mathcal{A} que se considera.

DEFINIÇÃO DE UMA MEDIDA FINITAMENTE ADITIVA. *Uma função de conjunto não negativa $f: \mathcal{A} \rightarrow \mathbf{R}$ que é finitamente aditiva diz-se que é uma medida finitamente aditiva, ou simplesmente uma medida.*

Aplicando o teorema 13.1 obtemos imediatamente as seguintes propriedades das medidas.

TEOREMA 13.2. *Se $f: \mathcal{A} \rightarrow \mathbf{R}$ é uma medida finitamente aditiva definida sobre uma álgebra de Boole \mathcal{A} , então para todos os conjuntos A e B de \mathcal{A} tem-se*

- (a) $f(A \cup B) \leq f(A) + f(B)$.
- (b) $f(B - A) = f(B) - f(A)$ se $A \subseteq B$.
- (c) $f(A) \leq f(B)$ se $A \subseteq B$. (*Propriedade monótona*)
- (d) $f(\emptyset) = 0$.

Demonstração. A alínea (a) resulta de (13.3), e a alínea (b) de (13.2). A alínea (c) é consequência de (b) e (d) obtém-se fazendo $A = B = \emptyset$ em (b).

EXEMPLO. O número de elementos de um conjunto finito. Seja $S = \{a_1, a_2, \dots, a_n\}$ um conjunto formado por n elementos (distintos), e seja \mathcal{A} a classe de todos os subconjuntos de S . Para cada A de \mathcal{A} , seja $\nu(A)$ o número de elementos distintos de A . É fácil verificar que esta função é finitamente aditiva em \mathcal{A} . Com efeito, se A tem k elementos e B tem m elementos, então $\nu(A) = k$ e $\nu(B) = m$. Se A e B são disjuntos é evidente que a união $A \cup B$ é um subconjunto de S com $k + m$ elementos, pelo que

$$\nu(A \cup B) = k + m = \nu(A) + \nu(B).$$

Esta função de conjunto particular é não negativa, pelo que ν é uma medida.

13.4. Exercícios

- Seja \mathcal{A} a classe de todos os subconjuntos de um dado conjunto fundamental e sejam A e B conjuntos arbitrários de \mathcal{A} . Provar que:
 - $A \cap B'$ e B são disjuntos.
 - $A \cup B = (A \cap B') \cup B$. (Esta igualdade exprime $A \cup B$ como a união de conjuntos disjuntos).
 - $A \cap B$ e $A \cap B'$ são disjuntos.
 - $(A \cap B) \cup (A \cap B') = A$. (Esta igualdade exprime A como uma união de dois conjuntos disjuntos.)
- O Exercício 1(b) evidencia a possibilidade de exprimir a união de dois conjuntos como uma união de dois conjuntos *disjuntos*. Exprimir de uma maneira análoga a união de três conjuntos $A_1 \cup A_2 \cup A_3$ e, mais geralmente, de n conjuntos $A_1 \cup A_2 \cup A_3 \dots \cup A_n$. Ilustrar com um diagrama para o caso em que $n = 3$.
- A análise de um conjunto S formado por 1000 licenciados universitários dez anos depois da licenciatura revelou que os "bem sucedidos" formavam um subconjunto A de 400 elementos, os licenciados por Caltech formavam um subconjunto B de 300 elementos, e que a intersecção $A \cap B$ era formada por 200 elementos.
 - Empregando a notação da teoria dos conjuntos, recorrendo às uniões e intersecções de A e B e dos seus complementos A' e B' relativamente a S , descrever os conjuntos de pessoas de S que gozam da propriedade seguinte:
 - Nem "bem sucedidos", nem licenciados por Caltech.
 - "Bem sucedidos", mas não licenciados por Caltech.
 - "Bem sucedidos", ou licenciados por Caltech, ou ambas as coisas.
 - Ou "bem sucedidos", ou licenciados por Caltech, mas não as duas coisas.
 - Pertencentes a não mais do que um dos subconjuntos A ou B .
 - Determinar o número exacto de indivíduos de cada um dos cinco subconjuntos precedentes.
- Seja f uma função de conjunto finitamente aditiva definida numa classe \mathcal{A} de conjuntos. Sejam A_1, \dots, A_n conjuntos de \mathcal{A} tais que $A_i \cap_m A_j = \emptyset$ se $i \neq j$. (Uma tal colecção diz-se uma *colecção disjunta* de conjuntos.) Se a união $\bigcup_{k=1}^m A_k$ pertence a \mathcal{A} para todo $m \leq n$, demonstrar por indução que

$$f\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n f(A_k).$$

Nos Exercícios 5, 6, 7 e 8, S representa um conjunto finito formado por n elementos distintos, a saber $S = \{a_1, a_2, \dots, a_n\}$.

5. Seja $A_1 = \{a_1\}$ o subconjunto formado unicamente pelo elemento a_1 . Provar que a classe $\mathcal{B}_1 = \{\emptyset, A_1, A_1^c, S\}$ é a mais pequena álgebra de Boole contendo A_1 .
6. Seja $A_1 = \{a_1\}$, $A_2 = \{a_2\}$. Descrever, de uma maneira semelhante à usada no Exercício 5, a menor álgebra de Boole \mathcal{B}_2 contendo quer A_1 , quer A_2 .
7. Resolver o exercício análogo ao anterior para os subconjuntos $A_1 = \{a_1\}$, $A_2 = \{a_2\}$, e $A_3 = \{a_3\}$.
8. Se \mathcal{B}_k representar a menor álgebra de Boole que contém os k subconjuntos $A_1 = \{a_1\}$, $A_2 = \{a_2\}$, ..., $A_k = \{a_k\}$, provar que \mathcal{B}_k contém 2^{k+1} subconjuntos de S se $k < n$ e 2^n subconjuntos se $k = n$.
9. Seja f uma função de conjunto finitamente aditiva, definida sobre a álgebra de Boole de todos os subconjuntos de um conjunto fundamental dado S . Admita-se que

$$f(A \cap B) = f(A)f(B)$$

para dois subconjuntos particulares A e B de S . Se $f(S) = 2$, provar que

$$f(A \cup B) = f(A') + f(B') - f(A')f(B').$$

10. Se A e B são dois conjuntos, a sua *diferença simétrica* $A \Delta B$ é o conjunto definido pela equação $A \Delta B = (A - B) \cup (B - A)$. Provar cada uma das seguintes propriedades da diferença simétrica.
 - (a) $A \Delta B = B \Delta A$.
 - (b) $A \Delta A = \emptyset$.
 - (c) $A \Delta B \subseteq (A \Delta C) \cup (C \Delta B)$.
 - (d) $A \Delta B$ é disjunto com cada um dos conjuntos A e B .
 - (e) $(A \Delta B) \Delta C = A \Delta (B \Delta C)$.
 - (f) Se f é uma função de conjunto finitamente aditiva sobre a álgebra de Boole \mathcal{A} de todos os subconjuntos de um lado conjunto S , então para todo o par A e B em \mathcal{A} tem-se $f(A \Delta B) = f(A) + f(B) - 2f(A \cap B)$.

13.5. A definição de probabilidade para conjuntos fundamentais finitos

Na linguagem das funções de conjunto, a probabilidade é um tipo específico de medida (aqui representada por P) definida sobre uma álgebra de Boole específica \mathcal{B} de conjuntos. Os elementos de \mathcal{B} são subconjuntos de um conjunto universal S . Na teoria das probabilidades o conjunto universal S chama-se o *conjunto fundamental*, *universo* ou *espaço amostra*. Vamos analisar a definição de probabilidade em primeiro lugar para conjuntos fundamentais finitos e mais tarde para espaços amostras infinitos.

DEFINIÇÃO DE PROBABILIDADE PARA CONJUNTOS FUNDAMENTAIS FINITOS. *Seja \mathcal{B} uma Álgebra de Boole cujos elementos são subconjuntos de um dado conjunto finito S .*

Uma função de conjunto P , definida sobre \mathcal{A} , diz-se uma medida de probabilidade se verifica as seguintes condições:

- (a) P é finitamente aditiva.
- (b) P é não negativa.
- (c) $P(S) = 1$.

Por outras palavras, isto significa que para conjuntos fundamentais finitos a probabilidade é simplesmente uma medida que atribui o valor 1 a todo o conjunto.

É importante ter presente que para uma descrição completa da medida de probabilidade devem precisar-se três coisas: O espaço amostra (ou conjunto fundamental) S , a álgebra de Boole \mathcal{B} formada por certos subconjuntos de S , e a função de conjunto P . O ternó (S, \mathcal{B}, P) designa-se frequentemente por *espaço de probabilidade*. Em muitas das aplicações elementares a álgebra de Boole \mathcal{A} é tomada como sendo a família de todos os subconjuntos de S .

EXEMPLO. O jogo de lançamento de uma moeda ao ar constitui uma aplicação prática da teoria das probabilidades. Para conjunto fundamental tomamos o conjunto de todos os resultados possíveis na experiência de lançamento da moeda ao ar. Neste caso, as hipóteses possíveis são “cara” ou “cruz”, que designaremos respectivamente por h e t . Assim, o conjunto fundamental S é $\{h, t\}$, o conjunto formado por h e t . Para álgebra de Boole tomamos a colecção de todos os subconjuntos de S ; são quatro no total, \emptyset , S , H e T , onde $H = \{h\}$ e $T = \{t\}$. Em seguida atribuímos probabilidades a cada um destes subconjuntos. Para os subconjuntos \emptyset e S não há opção para a escolha dos valores da probabilidade. A propriedade (c) requer que $P(S) = 1$, e, uma vez que P é uma medida não negativa, $P(\emptyset) = 0$. Todavia, há alguma liberdade na atribuição de probabilidades aos outros dois conjuntos, H e T . Visto que H e T são disjuntos e que a sua união é S , a propriedade aditiva exige que

$$P(H) + P(T) = P(S) = 1.$$

Para $P(H)$ e $P(T)$ podemos tomar valores quaisquer não negativos, e tais que a soma seja igual a 1. Se admitirmos que a moeda é imparcial não há *a priori* qualquer razão para preferir cara ou cruz, pelo que parece natural atribuir os valores

$$P(H) = P(T) = \frac{1}{2}.$$

Se, porém, a moeda estiver “viciada”, poderemos atribuir valores diferentes àqueles duas probabilidades. Por exemplo, os valores $P(H) = \frac{1}{3}$ e $P(T) = \frac{2}{3}$ são tão aceitáveis como $P(H) = P(T) = \frac{1}{2}$. Com efeito, para qualquer real p no intervalo $0 \leq p \leq 1$ podemos definir $P(H) = p$ e $P(T) = 1 - p$, e a função resultante P satisfará a todas as condições que se exigem para uma medida de probabilidade.

Para uma dada moeda, não existe um método matemático para determinar qual a probabilidade p que é a “real”. Se escolhermos $p = \frac{1}{2}$ podemos deduzir consequências lógicas na hipótese de que a moeda é perfeita ou imparcial. A teoria para moedas imparciais pode então ser usada para testar a perfeição de uma outra moeda, efectuan-

do um grande número de lançamentos e comparando os resultados com as predições baseadas na teoria. O testar o acordo entre a teoria e a evidência empírica pertence a um importante ramo das probabilidades aplicadas conhecido por *inferência estatística*, e não será estudado neste livro.

O exemplo precedente constitui uma aplicação típica dos conceitos da teoria das probabilidades. As questões probabilísticas têm muitas vezes lugar em situações designadas por "experiências". Não tentaremos definir o que se entende por uma experiência; em vez disso, mencionaremos alguns exemplos correntes: lançamento de uma ou mais moedas, lançamento de um par de dados, distribuir uma baralho de cartas, extrair uma bola de uma urna, contar o número de alunas de um determinado curso, selecção de um número numa lista telefónica, registo da radiação medida por um contador Geiger.

Para discutir as questões de probabilidade que estão ligadas a tais experiências, a nossa primeira tarefa é definir um conjunto fundamental S que possa usar-se para representar todas as hipóteses possíveis resultantes da experiência, como fizemos no lançamento da moeda ao ar. Cada elemento de S representará um resultado possível da experiência e cada resultado deverá corresponder a um e um só elemento de S . Em seguida, escolhemos uma álgebra de Boole \mathcal{B} de subconjuntos de S (habitualmente *todos* os subconjuntos de S) e então definimos uma medida de probabilidade P em \mathcal{B} . A escolha do conjunto S , a escolha de \mathcal{B} e a escolha de P dependerão da informação conhecida acerca dos detalhes da experiência e das questões a que pretendemos responder. O objectivo da teoria das probabilidades não é discutir se o espaço de probabilidade (S, \mathcal{B}, P) foi bem escolhido. Isto pertence à teoria dos jogos de azar, e apenas a expressão pode sugerir se sim ou não a escolha foi bem feita. *A teoria das probabilidades é o estudo das consequências lógicas que podem deduzir-se, uma vez definido o espaço de probabilidades.* Fazer uma boa escolha do espaço de probabilidade não é, estritamente, teoria de probabilidades—nem sequer é matemática; é, pelo contrário, uma parte da arte de aplicar a teoria das probabilidades ao mundo real. Nós aprofundaremos um pouco estas observações quando tratarmos com exemplos específicos nas secções que se seguem.

Se $S = \{a_1, a_2, \dots, a_n\}$, e se \mathcal{B} é formada por todos os subconjuntos de S , a função probabilidade P fica completamente determinada se conhecermos os seus valores nos subconjuntos formados por um só elemento.

$$P(\{a_1\}), P(\{a_2\}), \dots, P(\{a_n\}).$$

Com efeito, cada subconjunto A de S é uma união disjunta dos conjuntos anteriores, e $P(A)$ fica determinada pela propriedade aditiva. Por exemplo, quando

$$A = \{a_1\} \cup \{a_2\} \cup \dots \cup \{a_k\},$$

a propriedade aditiva exige que

$$P(A) = \sum_{i=1}^k P(\{a_i\}).$$

Para simplificar a notação e a terminologia, escrevemos $P(a_i)$ em vez de $P(\{a_i\})$. Este número chama-se também a *probabilidade do ponto* a_i . Consequentemente, a atribui-

ção de probabilidade pontuais $P(x)$ a cada elemento x de um conjunto finito S equivale a uma descrição completa da função de probabilidade P .

13.6. Terminologia peculiar da teoria das probabilidades

Ao falar-se de probabilidade escrevem-se muitas vezes frases como “dois acontecimentos são igualmente prováveis”, “um acontecimento é impossível”, ou “um acontecimento é certo”. Expressões desta natureza têm sentido intuitivo e é simultaneamente agradável e útil saber utilizar uma linguagem tão cheia de colorido nas discussões matemáticas. Antes, porém, de o podermos fazer torna-se necessário explicar o significado desta linguagem fazendo uso dos conceitos fundamentais da teoria.

Devido ao uso prático do ramo das probabilidades, é conveniente imaginar que cada espaço de probabilidade (S, \mathcal{B}, P) está associado a uma experiência real ou conceptual. O conjunto fundamental S pode então considerar-se como o conjunto de todos os resultados possíveis da experiência, tal como no exemplo do lançamento de uma moeda ao ar, já analisado na secção anterior. Cada elemento de S diz-se um *resultado* e os subconjuntos de S intervindo na álgebra booleana \mathcal{B} dizem-se *acontecimentos*. As razões desta terminologia tornar-se-ão mais evidentes quando tratarmos com exemplos concretos.

Suponhamos que temos um espaço de probabilidade (S, \mathcal{B}, P) associado a uma determinada experiência. Seja A um acontecimento e suponhamos a experiência realizada e que o resultado é x . (Por outras palavras, seja x um elemento de S). Este resultado x pode ou não pertencer ao conjunto A . Se pertence, dizemos que *ocorreu o acontecimento* A . Caso contrário, dizemos que *não ocorreu o acontecimento* A e então $x \in A'$, pelo que ocorreu o acontecimento complementar de A . Um acontecimento A diz-se *impossível* se $A = \emptyset$, porque neste caso nenhum resultado da experiência poderá pertencer a A . O acontecimento A diz-se ser o *acontecimento certo* se $A = S$, porque então todo o resultado é automaticamente um elemento de A .

Cada acontecimento A tem uma probabilidade $P(A)$ que lhe é definida pelo valor da função de probabilidade P . [O valor de $P(A)$ ou o modo segundo o qual $P(A)$ é definida não nos interessa por agora]. O número $P(A)$ chama-se também a *probabilidade de que o resultado de uma experiência seja um dos elementos de* A . Dizemos também que $P(A)$ é a *probabilidade de que o acontecimento* A *ocorra* quando se realiza a experiência.

Ao acontecimento impossível \emptyset deve atribuir-se a probabilidade zero porque P é uma medida finitamente aditiva. Todavia podem existir acontecimentos com probabilidade zero e que não são impossíveis. Quer dizer, a alguns dos subconjuntos não vazios de S pode atribuir-se a probabilidade zero. Ao acontecimento certo S atribui-se a probabilidade 1 segundo a correcta definição de probabilidade, mas podem existir outros subconjuntos aos quais seja atribuída a probabilidade 1. No Exemplo 1 da Secção 13.8 indicam-se subconjuntos não vazios com probabilidade zero e subconjuntos próprios de S que têm probabilidade 1.

Dois acontecimentos A e B dizem-se *igualmente prováveis* se $P(A) = P(B)$. O acontecimento A diz-se *mais provável* que B se $P(A) > P(B)$, e pelo menos *tão provável* como B se $P(A) \geq P(B)$. A tábuia 13.1 é um glossário da linguagem habitual nas discussões da

teoria das probabilidades. As letras A e B representam acontecimentos, e x o resultado de uma experiência associada ao conjunto fundamental S . Cada frase na coluna da esquerda é uma afirmação relativa aos acontecimentos A e B e a expressão correspondente na coluna da direita exprime a mesma afirmação em termos da teoria dos conjuntos

TÁBUA 13.1. Glossário de expressões usadas na teoria das probabilidades

<i>Afirmação</i>	<i>Significado na teoria dos conjuntos</i>
Pelo menos um dos acontecimentos A ou B ocorre	$x \in A \cup B$
Ocorrem ambos os acontecimentos A e B	$x \in A \cap B$
Nem A nem B ocorrem	$x \in A' \cap B'$
A ocorre e B não ocorre	$x \in A \cap B'$
Ocorre precisamente um dos acontecimentos A ou B	$x \in (A \cap B') \cup (A' \cap B)$
Não ocorre mais do que um dos acontecimentos A ou B	$x \in (A \cap B)'$
Se A ocorre, também ocorre B (A implica B)	$A \subseteq B$
A e B excluem-se mutuamente	$A \cap B = \emptyset$
Acontecimento A ou acontecimento B	$A \cup B$
Acontecimento A e acontecimento B	$A \cap B$

13.7. Exercícios

Seja S um dado espaço amostra e representem A , B e C acontecimentos arbitrários (isto é, subconjuntos de S na correspondente álgebra booleana \mathcal{B}). Cada uma das afirmações dos Exercícios 1 a 12 exprime-se por uma proposição verbal em termos de A , B e C . Exprimir estas proposições em termos de reuniões e intersecções de A , B e C e respectivos complementos.

1. Se A ocorre, então não ocorre B .
2. Nenhum dos acontecimentos A , B e C ocorre.
3. Ocorre unicamente A .
4. Pelo menos um dos acontecimentos A , B e C ocorre.
5. Ocorre precisamente um dos acontecimentos A , B e C .
6. Não ocorre mais do que um.
7. Pelo menos dois dos acontecimentos A , B , C ocorrem.
8. Ocorrem precisamente dois.
9. Ocorrem não mais do que dois.
10. Ocorrem A e C , mas não B .
11. Ocorrem os três acontecimentos.
12. Ocorrem não mais de três.
13. A designa o acontecimento de obter uma soma de pontos ímpar ao lançar dois dados, e B representa o acontecimento de obter pelo menos um seis. Exprimir, mediante uma proposição verbal, cada um dos acontecimentos seguintes:

- | | |
|-------------------|--------------------|
| (a) $A \cup B$, | (d) $A' \cap B$, |
| (b) $A \cap B$, | (e) $A' \cap B'$, |
| (c) $A \cap B'$, | (f) $A' \cup B$. |

14. Sejam A e B dois acontecimentos. Provar que

$$P(A \cap B) \leq P(A) \leq P(A \cup B) \leq P(A) + P(B).$$

15. Sejam A e B dois acontecimentos e designe-se $a = P(A)$, $b = P(B)$ e $c = P(A \cap B)$. Calcular, em função de a , b e c as probabilidades dos seguintes acontecimentos:

- | | |
|------------------|--------------------|
| (a) A' , | (d) $A' \cup B'$, |
| (b) B' , | (e) $A' \cup B$, |
| (c) $A \cup B$, | (f) $A \cap B'$. |

16. Dados três acontecimentos A , B e C provar que

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$

13.8. Exemplos resolvidos

Vamos mostrar como alguns conceitos das secções precedentes podem ser usados para responder a questões específicas da teoria das probabilidades.

EXEMPLO 1. Qual é a probabilidade de que ocorra pelo menos uma vez “cara” em dois lançamentos de uma moeda ao ar?

Primeira resolução. A experiência neste caso consiste do lançamento de uma moeda ao ar duas vezes; o conjunto S de todos os resultados possíveis pode exprimir-se do modo seguinte (h = cara, t = cruz).

$$S = \{hh, ht, th, tt\}.$$

Se aceitamos que estes resultados são igualmente prováveis, atribuímos a probabilidade pontual $P(x) = \frac{1}{4}$ para cada x de S . O acontecimento “ocorre a face cara pelo menos uma vez” pode representar-se pelo subconjunto

$$A = \{hh, ht, th\}.$$

A probabilidade deste acontecimento é a soma das probabilidades pontuais dos seus elementos. Por conseguinte, $P(A) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$:

Segunda resolução. Admitamos que se usa o mesmo espaço amostra, mas que atribuímos a probabilidade pontual do modo seguinte†:

$$P(hh) = 1, \quad P(ht) = P(th) = P(tt) = 0.$$

Então a probabilidade do acontecimento “ocorre pelo menos uma vez a face cara” é

† Observe-se que para esta atribuição de probabilidades existem subconjuntos não vazios de S com probabilidade zero e subconjuntos próprios com probabilidade 1.

$$P(hh) + P(ht) + P(th) = 1 + 0 + 0 = 1.$$

O facto de chegarmos a um resultado diferente do demonstrado no primeiro método de resolução não deve alarmar o leitor. Partimos de um conjunto de premissas diferentes. Considerações psicológicas podem levar-nos a acreditar que a atribuição das probabilidades na primeira solução é a mais natural. Na verdade, muita gente acreditará que é assim se a moeda for imparcial. Contudo, se a moeda estiver feita de modo que sempre saia cara, a atribuição de probabilidades na segunda solução é mais natural.

O exemplo precedente mostra que não podemos esperar uma única resposta para a pergunta feita. Para responder a uma tal questão de uma maneira adequada devemos especificar a escolha do conjunto fundamental e a atribuição das probabilidades pontuais. Uma vez que o conjunto fundamental e as probabilidades pontuais sejam conhecidas, apenas uma probabilidade pode ser logicamente deduzida para um dado acontecimento. Diferentes escolhas do espaço amostra ou das probabilidades pontuais podem conduzir a diferentes respostas "correctas" para a mesma questão.

Algumas vezes, a atribuição de probabilidades aos resultados particulares de uma experiência é ditada pela linguagem usada para descrever a experiência. Por exemplo, quando um objecto é escolhido "ao acaso" de um conjunto finito de n elementos, deve entender-se que com tal se pretende significar que cada resultado é igualmente provável e deverá atribuir-se-lhe a probabilidade pontual $1/n$. Analogamente, quando lançamos uma moeda ou um dado, se *a priori* não temos qualquer razão para pensar que a moeda ou o dado estão viciados, supomos que todos os resultados são igualmente prováveis. Adaptar-se-á este convénio em todos os exercícios deste capítulo.

EXEMPLO 2. Se extrairmos ao acaso duas cartas, uma de cada baralho, qual é a probabilidade de que uma, pelo menos, seja o ás de copas?

Resolução. A experiência consiste na tiragem de duas cartas a e b , uma de cada baralho. Suponhamos que representamos um resultado típico por um par ordenado (a, b) . O número de resultados possíveis, isto é, o número total de pares distintos (a, b) no espaço amostra S é 52^2 . Atribuimos a probabilidade $1/52^2$ a cada tal par. O acontecimento no qual estamos interessados que ocorra é o conjunto A dos pares (a, b) , nos quais ou a , ou b , é um ás de copas. Em A existem $52 + 51$ elementos. Logo, em face destas hipóteses concluimos que

$$P(A) = \frac{52 + 51}{52^2} = \frac{1}{26} - \frac{1}{52^2}.$$

EXEMPLO 3. Se de um baralho se extraem duas cartas ao acaso, qual é a probabilidade de que uma delas seja o ás de copas?

Resolução. Como no Exemplo 2, recorremos aos pares ordenados (a, b) para representar os elementos do espaço amostra. Neste caso o conjunto fundamental tem $52 \cdot 51$ elementos e o conjunto A tem $51 + 51$ elementos. Se atribuimos a probabilidade pontual $1/(52 \cdot 51)$ a cada resultado, obtemos

$$P(A) = \frac{2 \cdot 51}{52 \cdot 51} = \frac{1}{26}.$$

EXEMPLO 4. Na lançamento de três dados qual é a probabilidade de obter uma soma de pontos igual ou inferior a 6?

Resolução. Representamos todo o resultado possível da experiência por um terno de inteiros (a, b, c) , onde a, b, c podem tomar os valores de 1 até 6. Deste modo o conjunto fundamental é formado por 6^3 elementos e atribuímos a probabilidade $1/6^3$ a cada resultado. O acontecimento A em questão é o conjunto de todos os ternos verificando a desigualdade $3 \leq a + b + c \leq 6$. Se A_n representa o conjunto de (a, b, c) para o qual $a + b + c = n$, temos

$$A = A_3 \cup A_4 \cup A_5 \cup A_6.$$

Um cálculo directo mostra que os conjuntos A_n , com $n = 3, 4, 5$ e 6 contêm 1, 3, 6 e 10 elementos, respectivamente. Por exemplo, o conjunto A_6 é dado por

$$A_6 = \{(1, 2, 3), (1, 3, 2), (1, 1, 4), (1, 4, 1), (2, 1, 3), \\ (2, 3, 1), (2, 2, 2), (3, 1, 2), (3, 2, 1), (4, 1, 1)\}.$$

Deste modo A tem 20 elementos e

$$P(A) = \frac{20}{6^3} = \frac{5}{54}.$$

EXEMPLO 5. Lança-se um dado uma vez. Qual é a probabilidade de que o número de pontos obtido seja par ou múltiplo de 3?

Resolução. Escolhemos o espaço amostra $S = \{1, 2, 3, 4, 5, 6\}$, formado por seis elementos a cada um dos quais atribuímos a probabilidade $1/6$. O acontecimento "par" é o conjunto $A = \{2, 4, 6\}$, o acontecimento "múltiplo de 3" é $B = \{3, 6\}$. Interessa-nos a reunião destes conjuntos, a qual é o conjunto $A \cup B = \{2, 3, 4, 6\}$. Uma vez que este conjunto contém quatro elementos temos $P(A \cup B) = 4/6$.

Este exemplo pode resolver-se doutro modo, recorrendo-à fórmula

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{3}{6} + \frac{2}{6} - \frac{1}{6}.$$

13.9. Exercícios

1. Seja S um espaço amostra finito formado por n elementos. Admita-se que se atribui igual probabilidade a cada elemento de S . Seja A um subconjunto de S formado por k elementos. Provar que $P(A) = k/n$.

Em cada um dos Exercícios 2 a 8, descrever a escolha do espaço amostra e dizer como se atribuem as probabilidades pontuais. Nas questões associadas com jogos de cartas, supor que todos os cartas têm a mesma probabilidade de serem dadas extraídas do baralho.

2. Misturam-se cinco moedas falsas com nove moedas verdadeiras.
 - (a) Tira-se uma moeda ao acaso. Calcular a probabilidade de obter uma moeda falsa? Tiram-se duas moedas ao acaso; calcular a probabilidade de que:
 - (b) uma seja verdadeira e outra falsa.
 - (c) ambas sejam falsas.
 - (d) ambas sejam verdadeiras.
3. Calcular as probabilidades de cada um dos acontecimentos descritos no Exercício 13 da Secção 13.7. Atribuir probabilidades iguais a cada um dos 36 elementos do conjunto fundamental.
4. Qual a probabilidade de obter pelo menos uma das somas de pontos 7, 11 ou 12 no lançamento de dois dados?
5. Um jogador de poquer tem quatro copas e uma espada. Rejeita a espada e toma outra carta da parte restante do baralho. Calcular a probabilidade de tirar outra copa?
6. No poquer uma sequência é uma sucessão de cinco cartas, não necessariamente do mesmo naipe. Se um jogador de poquer possui quatro cartas em sucessão (mas não A234 ou JQKA) e uma quinta carta não em sucessão com as outras, calcular a probabilidade de completar a sequência. (O jogador desfaz-se da carta não em sucessão e tira nova carta da parte restante do baralho).
7. Um jogador de poquer tem quatro das cinco cartas em sucessão, mas com um salto no meio (como por exemplo, 5689) e uma quinta carta não em sucessão. Desfaz-se desta quinta carta e toma nova carta do resto do baralho. Calcular a probabilidade de completar a sequência.
8. Uma urna contém A bolas brancas e B bolas pretas. Uma segunda urna contém C bolas brancas e D bolas pretas. Tira-se uma bola ao acaso da primeira urna e introduz-se na outra. Tira-se depois ao acaso uma bola da segunda urna. Calcular a probabilidade de cada um dos seguintes acontecimentos.
 - (a) A primeira bola é branca.
 - (b) A primeira bola é preta.
 - (c) A segunda bola é branca, sabendo que era branca a bola introduzida na urna.
 - (d) A segunda bola é branca, sabendo que era preta a bola introduzida na urna.
9. De uma urna extraem-se duas bolas com reposição. A urna contém quatro bolas vermelhas e duas brancas. Calcular a probabilidade de cada um dos seguintes acontecimentos.
 - (a) Ambas as bolas são brancas.
 - (b) Ambas as bolas são vermelhas.
 - (c) Ambas as bolas são da mesma cor.
 - (d) Pelo menos uma das bolas é vermelha.
10. Seja P_n a probabilidade de que ocorram exactamente n dos acontecimentos A e B , onde n torna os valores 0, 1, 2. Expressar cada um dos valores P_0 , P_1 e P_2 em função de $P(A)$, $P(B)$, e $P(A \cap B)$.

Casos favoráveis. Alguns jogos de azar exprimem-se em termos de “casos favoráveis” em vez de probabilidades. Por exemplo, se lançamos um dado, a probabilidade do acontecimento “saída da face 3” é $\frac{1}{6}$. Uma vez que existem seis resultados possíveis, um dos quais é favorável e cinco desfavoráveis; muitas vezes exprimem-se tal facto dizendo que os casos a favor do acontecimento (saída da face três) estão como 1 par 5, ou que os casos contra o acontecimento é de 5 para 1. Costuma então relacionar-se isto com a probabilidade mediante a igualdade

$$\frac{1}{6} = \frac{1}{1+5}.$$

Em geral, se A é um acontecimento de probabilidade $P(A)$ e se a e b são dois números reais tais que

$$(13.6) \quad P(A) = \frac{a}{a+b},$$

dizemos que a *vantagem a favor* de A é de a para b , ou que a *vantagem contra* A é de b para a . Visto que $1 - a/(a+b) = b/(a+b)$, a sorte contra A é a mesma que a sorte a favor do acontecimento complementar A' .

Os exercícios seguintes destinam-se a evidenciar outras propriedades deste conceito dos "casos favoráveis" e suas relações com as probabilidades.

11. Se $P(A) = 1$, provar que (13.6) pode verificar-se somente quando $b = 0$ e $a \neq 0$. Se $P(A) \neq 1$, provar que há infinitas escolhas de a e b satisfazendo a (13.6), tendo porém todas elas o mesmo valor para o cociente a/b .
12. Calcular os casos favoráveis de cada um dos acontecimentos referidos no Exercício 2.
13. Dados dois acontecimentos A e B , se os casos contra são de 2 para 1 e os a favor de $A \cup B$ são de 3 para 1, provar que

$$\frac{5}{12} \leq P(B) \leq \frac{3}{4}.$$

Dar um exemplo no qual $P(B) = \frac{5}{12}$ e outro no qual $P(B) = \frac{3}{4}$.

13.10. Alguns princípios básicos de análise combinatória

Muitos problemas da teoria das probabilidades e de outros ramos da matemática podem reduzir-se a problemas de determinação do número de elementos de um conjunto finito. Os métodos sistemáticos para o estudo de tais problemas fazem parte de um ramo da matemática conhecido por *análise combinatória*. Nesta Secção vamos fazer uma breve digressão sobre este assunto para discutirmos alguns conceitos fundamentais que são úteis na análise de alguns problemas mais complicados da teoria das probabilidades.

Se todos os elementos de um conjunto finito são exibidos perante nós, não será difícil a contagem do total dos seus elementos. Muitas vezes acontece, porém, que um conjunto se define de tal maneira que é difícil, senão mesmo impossível, representar todos os seus elementos. Por exemplo, podemos perguntar qual o número total de conjuntos distintos de treze cartas que um jogador pode receber, quando, por quatro jogadores, é repartido um baralho de 52 cartas. Esse número supera os 635 mil milhões, e é evidente que uma enumeração directa de todos os casos possíveis não seria a melhor maneira de atacar o problema; contudo, recorrendo à análise combinatória a sua resolução é extremamente simples.

Este exemplo é um caso particular do problema mais geral de determinação do número de subconjuntos distintos de k elementos que podem formar-se a partir de um conjunto dado de n elementos†, com $n \geq k$. Representemos esse número por $f(n, k)$. Já se sabe que

† Quando dizemos que um conjunto tem n elementos queremos significar que esses n elementos são distintos.

$$(13.7) \quad f(n, k) = \binom{n}{k},$$

onde, como habitualmente, $\binom{n}{k}$ representa o coeficiente binomial,

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

No exemplo apresentado atrás temos $f(52, 13) = \binom{52}{13} = 635.013.559.600$ maneiras diferentes de distribuição das cartas a um jogador.

Conhecem-se vários métodos de demonstração de (13.7). O mais directo consiste em formar cada um dos subconjuntos de k elementos, pela escolha dos elementos um a um. Existem n possibilidades para a escolha do primeiro elemento, $n-1$ possibilidades para a escolha do segundo, e $n-(n-k-1)$ possibilidades para a escolha do k -énésimo. Fazendo todas as escolhas possíveis desta maneira obtemos um total de

$$n(n-1) \cdots (n-k+1) = \frac{n!}{(n-k)!}$$

subconjuntos de k elementos. É evidente que estes conjuntos não são todos distintos; por exemplo, se $k=3$ os seis subconjuntos

$$\{a, b, c\}, \{b, c, a\}, \{c, a, b\}, \{a, c, b\}, \{c, b, a\}, \{b, a, c\}$$

são todos iguais. Em geral, com este método cada subconjunto de k elementos conta-se precisamente $k!$ vezes.† Deste modo torna-se necessário dividir o número $n!/(n-k)!$ por $k!$ para obtermos $f(n, k)$. Temos pois que $f(n, k) = \binom{n}{k}$ como tínhamos afirmado.

Este tipo de raciocínio é mais ou menos característico da análise combinatória, por isso parece natural comentar ainda que resumidamente os fundamentos em que se baseia esta análise.

Muitas vezes pretendemos contar o número de elementos do produto cartesiano de n conjuntos finitos A_1, \dots, A_n . O produto cartesiano representa-se pelo símbolo $A_1 \times \cdots \times A_n$ e define-se pela igualdade

$$A_1 \times \cdots \times A_n = \{(a_1, \dots, a_n) \mid a_1 \in A_1, \dots, a_n \in A_n\}.$$

Quer dizer, o produto cartesiano é formado pelo conjunto de todos os n -tuplos ordenados (a_1, \dots, a_n) em que o elemento de ordem k provém do k -énésimo conjunto A_k .

Na figura 13.1 apresenta-se um exemplo com $n=2$. Aqui $A_1 = \{1, 2, 4, 5\}$ e $A_2 = \{1, 3\}$. Existem 4 elementos em A_1 , e 2 elementos em A_2 , dando um total de 8 elementos no produto cartesiano $A_1 \times A_2$. Mais geralmente, se A_1 é formado por k_1 elementos de A_2 formado por k_2 elementos, então $A_1 \times A_2$ é formado por $k_1 k_2$ ele-

† A razão de tal facto será esclarecida no exemplo 3 da pág. 543, onde se apresenta uma demonstração mais pormenorizada de (13.7).

mentos. Por indução em n , resulta que se A_r é formado de k_r elementos, então o produto cartesiano $A_1 \times \cdots \times A_n$ é formado por $k_1 \cdots k_n$ elementos.

Para se exprimir este resultado na terminologia das funções de conjunto, representamos por \mathcal{F} a classe de todos os conjuntos finitos e por ν a função de conjunto definida sobre \mathcal{F} do modo seguinte: Se $A \in \mathcal{F}$, $\nu(A)$ é o número de elementos distintos de A . (Para o conjunto vazio definimos $\nu(\emptyset) = 0$). Então é fácil verificar que ν é uma função de conjunto finitamente aditiva, pelo que se pode escrever

$$(13.8) \quad \nu\left(\bigcup_{i=1}^n S_i\right) = \sum_{i=1}^n \nu(S_i)$$

se $\{S_1, S_2, \dots, S_n\}$ é uma colecção de conjuntos finitos disjuntos (isto é, se $S_i \cap S_j = \emptyset$, sempre que $i \neq j$). O número de elementos de um produto cartesiano pode exprimir-se em função de ν do modo seguinte:

$$\nu(A_1 \times A_2 \times \cdots \times A_n) = \nu(A_1)\nu(A_2) \cdots \nu(A_n).$$

Uma fórmula parecida diz-nos como contar o número de elementos de qualquer conjunto T dos n -tuplos se soubermos o número das escolhas possíveis para cada uma das sucessivas componentes. Por exemplo, suponhamos que existem k_1 escolhas possíveis para a primeira componente x_1 . Seja k_2 o número das escolhas possíveis para a segunda componente x_2 , uma vez escolhido x_1 . Analogamente seja k_r o número das escolhas possíveis para a componente x_r , uma vez escolhidos x_1, x_2, \dots, x_{r-1} . Então o número dos n -tuplos que podem formar-se com estas escolhas é

$$\nu(T) = k_1 k_2 \cdots k_n.$$

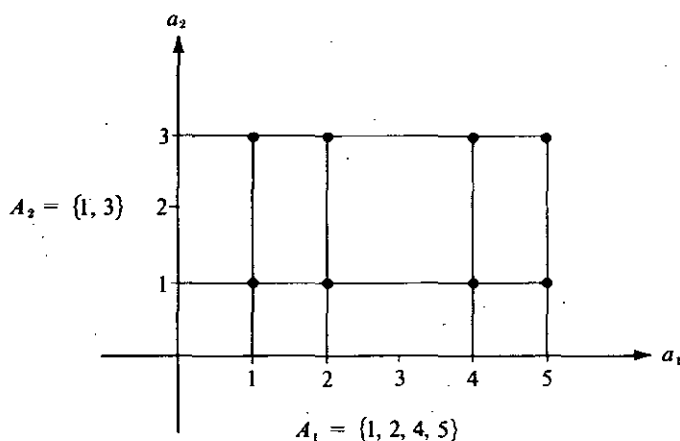


FIG. 13.1. Um exemplo ilustrando o produto cartesiano de dois conjuntos. Os pontos marcados representam $A_1 \times A_2$.

Esta fórmula designa-se muitas vezes por *regra da contagem sucessiva*. Pode ser demonstrada por indução em n . Em muitas aplicações o conjunto das escolhas para x_r pode não ser fácil de descrever, pois que não pode ser determinado senão depois de ter sido feita a escolha dos elementos anteriores. Felizmente, para aplicar a regra da contagem sucessiva não necessitamos conhecer o conjunto actual das escolhas para x_r , mas *unicamente o número* das escolhas possíveis para x_r .

A propriedade aditiva na fórmula (13.8) e a regra da contagem sucessiva proporcionam-nos a solução de muitos problemas de cálculo. Nos exemplos seguintes vamos analisar o modo de os aplicar.

EXEMPLO 1. Tiragens com reposição. Dado um conjunto S formado por n elementos, se $k \geq 1$ quantos k -tuplos ordenados podem formar-se se cada componente pode ser um elemento qualquer de S ?

Nota: Pode ser útil imaginar S como uma urna contendo n bolas numeradas de 1 até n . Tiramos uma bola e consideramos o seu número como o primeiro elemento do nosso k -tuplo. Colocamos de novo a bola na urna, e escolhemos de novo uma bola e anotamos o seu número como segunda componente e assim sucessivamente, até que tenhamos feito k tiragens. Uma vez que a tiragem de cada bola é feita com reposição, o mesmo número pode aparecer em diferentes componentes do nosso k -tuplo.

Resolução. Cada k -tuplo é um elemento do produto cartesiano

$$T = S_1 \times \cdots \times S_k,$$

onde cada $S_i = S$. Reciprocamente, cada elemento de T é um dos k -tuplos em questão. Logo o número dos k -tuplos formados deste modo é

$$v(T) = v(S_1) \cdots v(S_k) = n^k.$$

EXEMPLO 2. Tiragens sem reposição. Dado um conjunto S formado por n elementos, se $k \leq n$, quantos k -tuplos ordenados se podem formar se as componentes são escolhidas de S *sem* reposição, quer dizer, se nenhum elemento de S pode aparecer repetindo um mesmo k -tuplo?

Resolução. Consideremos qualquer k -tuplo (x_1, x_2, \dots, x_k) formado a partir dos elementos de S *sem* reposição. Para o primeiro elemento x_1 temos n possibilidades de escolha (os n elementos de S). Uma vez escolhido x_1 , restam $n - 1$ maneiras de escolher x_2 . Com x_2 escolhido, restam $n - 2$ maneiras de escolher x_3 , e assim sucessivamente, sendo $n - k + 1$ as possibilidades de escolha de x_k . Consequentemente, pela regra da contagem sucessiva, o número total dos k -tuplos assim formados é

$$n(n-1)(n-2) \cdots (n-k+1) = \frac{n!}{(n-k)!}.$$

Em particular, quando $k = n$ este resultado diz-nos que de um conjunto de n elementos se podem formar $n!$ n -tuplos distintos, sem que qualquer deles contenha elementos repetidos.

EXEMPLO 3. *O número de subconjuntos de k elementos de um conjunto de n elementos. Se $k \leq n$, quantos subconjuntos distintos de k elementos se podem formar com os n elementos do conjunto S ?*

Resolução. Seja r o número de subconjuntos em questão e representemos esses subconjuntos por

$$A_1, A_2, \dots, A_r.$$

Estes conjuntos são distintos, mas não necessariamente disjuntos. Podemos determinar r em função de n e k por via indirecta. Com tal finalidade, representemos por B_i a colecção dos k -tuplos que podem formar-se pela escolha dos componentes a partir dos elementos de A_i sem reposição. Os conjuntos B_1, B_2, \dots, B_r são disjuntos. Além disso, se aplicarmos o resultado do Exemplo 2 com $n = k$ temos

$$\nu(B_i) = k! \quad \text{para todo } i = 1, 2, \dots, r.$$

Façamos agora

$$T = B_1 \cup B_2 \cup \dots \cup B_r.$$

Então T consta de todos os k -tuplos que podem formar-se escolhendo as componentes em S sem reposição. Do Exemplo 2 resulta

$$\nu(T) = n!/(n - k)!$$

e pela aditividade temos também

$$\nu(T) = \sum_{i=1}^r \nu(B_i) = k! r.$$

Igualando as duas expressões de $\nu(T)$ obtemos

$$r = \frac{n!}{k!(n - k)!} = \binom{n}{k}.$$

Isto prova a fórmula (13.7), já estabelecida atrás nesta Secção.

Se usarmos o resultado do Exemplo 3 para calcular o número total de subconjuntos de um conjunto S formado por n elementos, obtemos

$$\sum_{k=0}^n \binom{n}{k}.$$

Uma vez que esta soma se obtém igualmente no desenvolvimento de $(1 + 1)^n$ pela fórmula do binómio, o número de subconjuntos de S é 2^n .

13.11. Exercícios

- Seja $A = \{1, 2, 3\}$. Representar em extensão o conjunto dos pares ordenados (a, b) obtidos por escolha da primeira componente em A e a segunda componente entre os *restantes* elementos de A . Pode este conjunto de pares ordenados exprimir-se na forma de um produto cartesiano?
- De um baralho de 52 cartas podem tirar-se duas cartas de $52 \cdot 51 = 2652$ maneiras. Determinar o número dos pares *distintos* de cartas e expôr o raciocínio.
- Um comité do senado norte americano formado por 6 democratas e 4 republicanos tem que eleger um presidente e um vice presidente. De quantas maneiras se pode escolher este par de dirigentes se o presidente dever ser um democrata?
- Um jogo consiste no lançamento ao ar de uma moeda duas vezes, seguido do lançamento de um dado. Descrever o resultado deste jogo como um terno ordenado (a, b, c) , em que a e b podem ser ou "cara" (H) ou "cruz" (T) e c é o número de pontos marcados no dado. Por exemplo $(H, H, 3)$ significa que em ambos os lançamentos da moeda se obteve a face cara e o dado marcou três pontos. Expressar o conjunto de todos os resultados possíveis por um produto cartesiano e determinar o número de resultados possíveis.
- De quantas maneiras possíveis pode distribuir-se por quatro jogadores um baralho de 52 cartas, ficando cada um com 13 cartas? Explicar o raciocínio efectuado.
- Lançam-se simultaneamente dois dados, um vermelho e outro preto. Representar o resultado por um par ordenado (a, b) , em que a representa o número de pontos do dado vermelho e b os do dado preto. Quantos pares ordenados (a, b) são possíveis? Quantos existem para os quais a soma $a + b$ é:
(a) par? (b) divisível por 3? (c) ou par ou divisível por 3?
- Uma mão de poker é formada por 5 cartas extraídas de um baralho de 52. Quantas mãos distintas podem formar-se contendo:
(a) dois pares (por exemplo, 2 reis, 2 duques e 1 terno)?
(b) cinco cartas do mesmo naipe?
(c) cinco cartas do mesmo naipe consecutivas, não incluindo o dez, valete, dama, rei e às?
(d) dez, valete, dama, rei e às do mesmo naipe?
- Considerar o Exercício 7. Calcular a probabilidade para que uma mão de poker seja:
(a) cinco cartas do mesmo naipe.
(b) cinco cartas do mesmo naipe seguidas (não incluindo o dez, valete, dama, rei e às).
(c) dez, valete, dama, rei e às do mesmo naipe.
- Quantos comités de 50 senadores se podem formar de modo a conterem:
(a) precisamente um só senador do Alasca?
(b) ambos os senadores do Alasca?
- Escolhe-se ao acaso um comité de 50 senadores. Calcular a probabilidade de que a ele pertençam os dois senadores do Alasca?
- Forma-se um código com quatro símbolos em linha, sendo cada símbolo ou um ponto ou um traço. Quantos grupos distintos do código podem formar-se?
- Quantas palavras de k letras se podem formar com um alfabeto de n letras?

13. Provar que

$$(a) \binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \cdots = \binom{n}{1} + \binom{n}{3} + \binom{n}{5} + \cdots = 2^{n-1}.$$

$$(b) \binom{n}{0}^2 + \binom{n}{1}^2 + \cdots + \binom{n}{n}^2 = \binom{2n}{n}.$$

14. Supor um conjunto de pares ordenados (a, b) construído pela escolha da primeira componente entre os k elementos de um conjunto, por exemplo $\{a_1, \dots, a_k\}$ e a segunda componente entre os m elementos do conjunto $\{b_1, \dots, b_m\}$. Existem m pares com a primeira componente a_1 , nomeadamente, $(a_1, b_1), (a_1, b_2), \dots, (a_1, b_m)$. Analogamente existem m pares $(a_i, b_1), \dots, (a_i, b_m)$ tendo por primeira componente o elemento a_i . Portanto o número total de pares ordenados (a, b) é $m + m + \cdots + m$ (k vezes). Esta soma vale km , o que prova a regra de contagem sucessiva para conjuntos de pares ordenados. Utilizar o método de indução para provar essa regra para conjuntos de n -tuplos ordenados.

13.12. Probabilidade condicionada

Lança-se um dado perfeito e sabe-se que o número de pontos é par. Qual a probabilidade de que esse número seja divisível por 3? Qual a probabilidade de uma criança seja daltônica, sabendo que é um rapaz? Estas questões podem formular-se do modo seguinte: Sejam A e B acontecimentos de um conjunto fundamental S . Qual é a probabilidade de que A ocorra, sabido que B ocorreu? Isto não é necessariamente o mesmo que pedir a probabilidade do acontecimento $A \cap B$. Com efeito, quando $A = B$ a questão põe-se do modo seguinte: Se A ocorre, qual é a probabilidade de que A ocorra? A resposta deve ser evidentemente 1 e esta pode ser, ou não, a probabilidade de $A \cap B$. Para analisarmos tais problemas em geral, voltamos ao caso do dado já referido.

Quando tratamos de questões de probabilidades relativas ao lançamento de um dado imparcial, consideramos geralmente como conjunto fundamental o conjunto $S = \{1, 2, 3, 4, 5, 6\}$ e atribuímos a probabilidade $\frac{1}{6}$ a cada elemento de S . O acontecimento "divisível por 3" é o subconjunto $A = \{3, 6\}$ e o acontecimento "par" é o subconjunto $B = \{2, 4, 6\}$. Pretendemos a probabilidade de que um elemento esteja em A , sabendo que pertence a B . Visto que estamos unicamente interessados em resultados nos quais o número é par, pomos de parte os resultados 1, 3, 5 e usamos como conjunto fundamental em vez de S , o conjunto $B = \{2, 4, 6\}$. O acontecimento no qual estamos agora interessados é unicamente o conjunto de um elemento, $\{6\}$, sendo este o único elemento do conjunto fundamental que é divisível por 3. Se todos os resultados de B são igualmente prováveis, devemos atribuir a probabilidade $\frac{1}{3}$ a cada um deles; por isso, em particular, a probabilidade de $\{6\}$ é também $\frac{1}{3}$.

Observe-se que resolvemos o problema com recurso a uma ideia muito elementar. Muito simplesmente mudámos o espaço amostra de S para B e procedemos a uma nova atribuição de probabilidades. Isto sugere-nos a maneira de generalizar o processo.

Seja (S, \mathcal{B}, P) um dado espaço de probabilidade. Sejam A e B dois acontecimentos e consideremos a questão: "Qual é a probabilidade de que A ocorra, sabido que B ocorreu"? Como no exemplo acabado de considerar, podemos mudar o conjunto

fundamental de S para B e proceder a nova atribuição de probabilidades. Temos a liberdade de fazer isto de qualquer maneira consistente com a definição de medidas de probabilidade. Para B não temos escolha, excepto para atribuir a probabilidade 1. Uma vez que estamos interessados naqueles elementos de A que pertencem ao novo conjunto fundamental B , o problema é o de calcular a probabilidade do acontecimento $A \cap B$, de acordo com a nova distribuição de probabilidades. Quer dizer, se P' representa a função de probabilidade associada com o novo espaço amostra B , então temos que calcular $P'(A \cap B)$.

Demonstraremos agora que se $P(B) \neq 0$ podemos sempre definir uma função probabilidade P' e uma álgebra de Boole; \mathcal{B}' , de subconjuntos de B , de maneira que $C(B, \mathcal{B}', P')$ seja um espaço de probabilidade. Para a álgebra de Boole \mathcal{B}' tomamos a família de todos os conjuntos $T \cap B$ em que T é um conjunto que pertence à álgebra de Boole original \mathcal{B} . É fácil verificar que \mathcal{B}' , assim definida, é na verdade uma álgebra de Boole. Uma maneira de definir uma função de probabilidade P' sobre \mathcal{B}' consiste em dividir cada uma das primitivas probabilidades por $P(B)$. Quer dizer, se $C \in \mathcal{B}'$ temos

$$P'(C) = \frac{P(C)}{P(B)}.$$

(Isto evidentemente na hipótese de $P(B) \neq 0$.) O que se fez foi unicamente uma mudança de escala, com todas as probabilidades multiplicadas por um factor $1/P(B)$. É fácil verificar que esta definição de P' nos dá uma *bona fide* medida de probabilidade. É evidentemente não negativa e atribui a probabilidade 1 a B . A propriedade aditiva resulta, por sua vez, da correspondente propriedade aditiva de P .

Visto cada C em \mathcal{B}' ser da forma $A \cap B$, em que A é um acontecimento no espaço amostra original S , podemos de novo escrever a definição de P' na forma:

$$P'(A \cap B) = \frac{P(A \cap B)}{P(B)}.$$

Esta discussão sugere que o cociente $P(A \cap B)/P(B)$ proporciona uma medida de probabilidade conveniente de que A ocorra, sabido que B ocorreu. A definição que apresentamos seguir é dada tendo em mente o que se acabou de expôr:

DEFINIÇÃO DE PROBABILIDADE CONDICIONADA. *Sejam (S, \mathcal{B}, P) um espaço de probabilidade e B um acontecimento tal que $P(B) \neq 0$. A probabilidade condicionada de que um acontecimento A ocorra, na hipótese de que B ocorreu, representa-se pelo símbolo $P(A | B)$ (lê-se: "a probabilidade de A , suposto realizado B ") e define-se pela igualdade*

$$P(A | B) = \frac{P(A \cap B)}{P(B)}.$$

A probabilidade condicionada $P(A | B)$ não fica definida se $P(B) = 0$.

Os exemplos que se seguem ilustram o uso deste conceito de probabilidade condicionada.

EXEMPLO 1. Considere-se, uma vez mais, o problema posto no início desta Secção: Lança-se um dado e sabe-se que o número de pontos obtido é par. Qual a probabilidade de que seja divisível por 3? Como problema de probabilidade condicionada, podemos tomar para conjunto fundamental o conjunto $S = \{1, 2, 3, 4, 5, 6\}$ e atribuir a cada elemento a probabilidade $\frac{1}{6}$. O acontecimento "par" é o conjunto $B = \{2, 4, 6\}$ e o acontecimento "divisível por 3" é o conjunto $A = \{3, 6\}$. Deste modo temos

$$P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{3/6} = \frac{1}{3}.$$

o que está, evidentemente, de acordo com a anterior solução na qual considerámos B como espaço amostra e atribuímos a probabilidade $\frac{1}{3}$ a cada um dos seus elementos.

EXEMPLO 2. Trata-se de um exemplo típico usado no Departamento de Biologia de Caltech para prevenir contra os erros de estatísticas superficiais. Para "provar" estatisticamente que a população dos Estados Unidos da América contém mais rapazes do que raparigas, pede-se a cada estudante para indicar o número de rapazes e raparigas da sua família. Invariavelmente, o número total de rapazes excede o número total de raparigas. As estatísticas neste caso estão viciadas porque todos os estudantes de Caltech são rapazes. Consequentemente, a questão que aqui se considera não está relacionada com a probabilidade de que uma indivíduo seja um rapaz; antes pelo contrário, está relacionada com a probabilidade condicionada de que um indivíduo seja rapaz, sabido que pertence a uma família onde existe pelo menos um rapaz.

Para se calcularem as probabilidades num exemplo deste tipo considera-se uma amostra de $4n$ famílias, cada uma com dois filhos. Admita-se que n famílias têm 2 rapazes e $2n$ famílias têm um rapaz e uma rapariga, e n famílias 2 raparigas. Seja o espaço amostra S o conjunto de todos os $8n$ filhos destas $4n$ famílias e atribua-se a probabilidade $P(x) = 1/(8n)$ a cada x de S . Represente A o acontecimento "o indivíduo é um rapaz" e B o acontecimento "o indivíduo procede de uma família com pelo menos um rapaz". A probabilidade $P(A)$ é evidentemente $\frac{1}{2}$. Analogamente $P(B) = \frac{3}{4}$, uma vez que $3n$ das $4n$ famílias têm pelo menos um rapaz. Portanto a probabilidade de que o indivíduo seja um rapaz, sabido que é oriundo de uma família com pelo menos um rapaz, é a probabilidade condicionada.

$$P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{1/2}{3/4} = \frac{2}{3}.$$

13.13. Independência aleatória

Um conceito importante relacionado com a probabilidade condicionada é o de *independência de acontecimentos*, que se pode definir do modo seguinte:

DEFINIÇÃO DE INDEPENDÊNCIA ALEATÓRIA. *Dois acontecimentos A e B dizem-se independentes (ou aleatoriamente independentes) se e só se*

$$(13.9) \quad P(A \cap B) = P(A)P(B).$$

Se A e B são independentes, então $P(A|B) = P(A)$ se $P(B) \neq 0$, isto é, a probabilidade condicionada de A , dado B , é a mesma que a probabilidade “absoluta” de A . Esta relação põe em evidência o significado de independência. O conhecimento de que B ocorreu não influencia a probabilidade de que A ocorra.

EXEMPLO 1. Faz-se uma tiragem à sorte de uma carta de um baralho de 52 cartas. Todas as cartas têm a mesma probabilidade de serem escolhidas. Demonstrar que os dois acontecimentos “tirar um ás” e “tirar uma copa” são independentes.

Resolução. Definimos um conjunto fundamental S de 52 elementos e atribuímos a cada elemento a probabilidade $\frac{1}{52}$. O acontecimento A “tirar um ás” tem a probabilidade $P(A) = \frac{4}{52} = \frac{1}{13}$. O acontecimento B , “tirar uma copa” tem a probabilidade $P(B) = \frac{13}{52} = \frac{1}{4}$. O acontecimento $A \cap B$ significa “tirar o ás de copas”, o qual tem a probabilidade $\frac{1}{52}$. Uma vez que $P(A \cap B) = P(A) \cdot P(B)$, a igualdade (13.9) garante-nos que os acontecimentos A e B são independentes.

EXEMPLO 2. Lançam-se independentemente três dados “perfeitos”, de modo que toda a combinação de pontos seja igualmente provável. Seja A o acontecimento “soma dos pontos obtidos igual a 6” e B o acontecimento “os pontos indicados pelos três dados são diferentes”. Determinar se sim ou não estes dois acontecimentos são independentes.

Resolução. Para conjunto fundamental S tomamos todos os ternos de valores (a, b, c) , podendo a, b e c tomarem os valores 1, 2, 3, 4, 5, 6. S contém 6^3 elementos e porque são igualmente prováveis terá cada um a probabilidade $1/6^3$. O acontecimento A é o conjunto de todos os ternos (a, b, c) para os quais $a + b + c = 6$. Uma enumeração directa desses casos mostra que eles são em número de dez, a saber:

(1, 2, 3), (1, 3, 2), (1, 1, 4), (1, 4, 1),

(2, 1, 3), (2, 3, 1), (2, 2, 2),

(3, 1, 2), (3, 2, 1),

(4, 1, 1).

O acontecimento B consiste de todos os ternos (a, b, c) para os quais $a \neq b, b \neq c$ e $a \neq c$. Existem $6 \cdot 5 \cdot 4 = 120$ elementos em B . Precisamente 6 destes elementos pertencem ao conjunto A , pelo que $A \cap B$ tem seis elementos. Deste modo

$$P(A \cap B) = \frac{6}{6^3}, \quad P(A) = \frac{10}{6^3}, \quad \text{e} \quad P(B) = \frac{120}{6^3}.$$

Neste caso $P(A \cap B) \neq P(A)P(B)$; portanto os acontecimentos A e B não são independentes.

A independência para mais do que dois acontecimentos define-se do modo seguinte: uma família \mathcal{A} de um número finito n de acontecimentos diz-se independente se os acontecimentos verificam a propriedade

$$(13.10) \quad P\left(\bigcap_{k=1}^m A_k\right) = \prod_{k=1}^m P(A_k)$$

para *todo* o subconjunto $\{A_1, A_2, \dots, A_m\}$, com m a tomar os valores $m = 2, 3, \dots, n$ e os conjuntos A_i pertencentes a \mathcal{A} .

Quando \mathcal{A} é precisamente formado por três acontecimentos A , B e C , a condição de independência (13.10) requer que

$$(13.11) \quad P(A \cap B) = P(A)P(B), \quad P(A \cap C) = P(A)P(C), \quad P(B \cap C) = P(B)P(C),$$

e

$$(13.12) \quad P(A \cap B \cap C) = P(A)P(B)P(C).$$

Pode pensar-se que as três equações (13.11) são suficientes para que se verifique (13.12) ou, por outras palavras, que a independência de três acontecimentos é uma consequência da independência desses acontecimentos, considerados *dois a dois*. Isto porém, não é verdadeiro, como se pode ver através do exemplo seguinte:

Quatro cartões marcados com a , b , c e abc , colocam-se dentro de uma caixa. Tira-se um cartão ao acaso, e o conjunto fundamental representa-se por

$$S = \{a, b, c, abc\}.$$

Definem-se os acontecimentos A , B , C do modo seguinte:

$$A = \{a, abc\}, \quad B = \{b, abc\}, \quad C = \{c, abc\}.$$

Quer dizer, o acontecimento X significa que o cartão extraído contém a letra x . É fácil verificar que cada uma das três igualdades (13.11) é satisfeita, pelo que os três acontecimentos A , B , C são independentes dois a dois. Contudo (13.12) não se verifica e por conseguinte os *três* acontecimentos não são independentes. Os cálculos são simples e deixam-se ao leitor como exercício.

13.14 Exercícios

1. Sejam A e B dois acontecimentos com $P(A) \neq 0$, $P(B) \neq 0$, Provar que

$$(13.13) \quad P(A \cap B) = P(B)P(A|B) = P(A)P(B|A).$$

Algumas vezes é mais fácil calcular as probabilidades $P(A)$ e $P(B|A)$ directamente, do que $P(A \cap B)$. Quando tal acontece, a equação (13.13) dá-nos um modo conveniente de calcular $P(A \cap B)$. O exercício que se segue é disso exemplo.

2. Uma urna contém sete bolas brancas e três pretas. Uma segunda urna contém cinco bolas brancas e cinco pretas. Extrai-se uma bola ao acaso da primeira urna e introduz-se na segunda. Extrai-se em seguida uma bola ao acaso da segunda urna. Seja A o acontecimento "bola preta na primeira extracção" e B o acontecimento "bola preta na segunda extracção":
- (a) Calcular as probabilidades $P(A)$ e $P(B|A)$ directamente, por enumeração dos casos possíveis. Recorrer a (13.13) para calcular $P(A \cap B)$.
- (b) Calcular $P(A \cap B)$ directamente por enumeração de todos os pares de extracções possíveis.
3. (a) Sejam A_1, A_2 e A_3 três acontecimentos tais que $P(A_1 \cap A_2) \neq 0$. Provar que

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2).$$

(b) Utilizar o método de indução para generalizar este resultado do modo seguinte: Se A_1, A_2, \dots, A_n são n ($n \geq 2$) acontecimentos tais que $P(A_1 \cap A_2 \cap \dots \cap A_{n-1}) \neq 0$, então

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1}).$$

4. Escolhe-se ao acaso um comité de 50 senadores. Determinar a probabilidade de que ambos os senadores do Alasca estejam incluídos entre eles, sabido que pelo menos um deles o está.
5. Uma urna contém cinco bolas douradas e sete bolas azuis. Extraem-se ao acaso duas bolas (sem reposição). Se a primeira for dourada, calcular a probabilidade de que a segunda o seja também.
6. Um baralho de 52 cartas é repartido por quatro jogadores em partes iguais. Se um deles tem sete espadas, qual é a probabilidade de que um outro determinado jogador tenha (a) pelo menos uma espada? (b) pelo menos duas espadas? (c) um naipe completo?
7. Provar que $P(A \cup B|C) = P(A|C) + P(B|C) - P(A \cap B|C)$.
8. Sejam A_1, A_2, \dots, A_n n acontecimentos disjuntos cuja reunião é o espaço amostra S . Para cada acontecimento E tem-se a equação

$$E = E \cap S = E \cap \bigcup_{i=1}^n A_i = \bigcup_{i=1}^n (E \cap A_i).$$

Esta equação estabelece que E pode ocorrer unicamente em conjunção com algum A_i . Demonstrar que

$$(a) P(E) = \sum_{i=1}^n P(E \cap A_i).$$

$$(b) P(E) = \sum_{i=1}^n P(E|A_i)P(A_i).$$

Esta fórmula é útil quando as probabilidades condicionadas $P(E|A_i)$ são mais fáceis de calcular directamente do que $P(E)$.

9. Uma moeda "imparcial" lança-se ao ar repetidas vezes. Nos seis primeiros lançamentos sai sempre a face cara. Qual a probabilidade de que sai cara no lançamento seguinte?
10. São dados dois acontecimentos independentes A e B cujas probabilidades de ocorrência não são nem 0, nem 1. Demonstrar que A' e B' são independentes. O mesmo será verdadeiro caso A ou B tenham probabilidades 0 ou 1?

11. A e B são dois acontecimentos independentes. Provar as alíneas seguintes ou, se for o caso, provar que são falsas.
- A' e B são independentes.
 - $A \cup B$ e $A \cap B$ são independentes.
 - $P(A \cup B) = 1 - P(A')P(B')$.
12. Se A_1, A_2, \dots, A_n são acontecimentos independentes, provar que

$$P\left(\bigcup_{i=1}^n A_i\right) + \prod_{i=1}^n P(A_i') = 1.$$

13. Se três acontecimentos A, B e C são independentes, provar que $A \cup B$ e C são independentes. [Sugestão: Utilizar o resultado do Exercício 7 para provar que $P(A \cup B | C) = P(A \cup B)$].
14. Sejam A e B dois acontecimentos, nenhum deles com probabilidade 0. Provar as alíneas seguintes ou, se for o caso, provar que são falsas:
- Se A e B são disjuntos, A e B são independentes.
 - Se A e B são independentes, A e B são disjuntos.
15. Lança-se um dado duas vezes, sendo o conjunto fundamental formado pelos 36 pares possíveis de resultados (a, b) , cada um dos quais com probabilidade $\frac{1}{36}$. Sejam A, B e C os seguintes acontecimentos:

$$A = \{(a, b) \mid a \text{ é ímp}\}, \quad B = \{(a, b) \mid b \text{ é ímp}\}, \quad C = \{(a, b) \mid a + b \text{ é ímp}\}.$$

- Calcular $P(A), P(B), P(C), P(A \cap B), P(A \cap C), P(B \cap C)$ e $P(A \cap B \cap C)$.
- Provar que A, B e C são independentes dois a dois.
- Provar que A, B e C não são independentes.

13.15 Experiências compostas

Voltamos de novo a nossa atenção para o problema de de Méré referido na introdução-se será ou não conveniente apostar a mesma importância a favor ou contra a obtenção de pelo menos um “duplo seis” em 24 lançamentos de um par de dados. Vamos analisar o problema de uma forma geral: qual a probabilidade de obter um “duplo seis”, pelo menos uma vez em n lançamentos de um par de dados? Será esta probabilidade superior ou inferior a $1/2$ quando $n = 24$?

Consideremos em primeiro lugar a experiência que consiste no lançamento de um par de dados numa só vez. Os resultados possíveis podem esquematizar-se por pares ordenados (a, b) , nos quais a e b tomam os valores 1, 2, 3, 4, 5, 6. O conjunto fundamental S é formado por 36 pares desse tipo. Uma vez que os dados são “perfeitos” atribuímos a probabilidade $\frac{1}{36}$ a cada elemento de S .

Suponhamos agora que efectuamos n lançamentos. A sucessão das n experiências definirá uma experiência composta que desjamos descrever matematicamente. Para o conseguirmos necessitamos definir um novo conjunto fundamental e a correspondente medida de probabilidade. Consideramos os resultados do novo jogo com n -tuplos ordenados (x_1, x_2, \dots, x_n) onde cada componente x_i é um dos resultados do espaço amostra inicial S . Por outras palavras, o espaço amostra para a experiência composta é o produto cartesiano $S \times \dots \times S$ que representamos por S^n . O conjunto S^n tem 36^n elementos, e a cada elemento atribuímos a probabilidade

$$P(x) = \frac{1}{36^n} \quad \text{se } x \in S^n.$$

Estamos interessados na realização do acontecimento “pelo menos uma vez um duplo seis em n lançamentos”. Representamos este acontecimento por A . Neste caso é mais fácil calcular a probabilidade do acontecimento complementar A' que significa “nenhum duplo seis em n lançamentos”. Cada elemento de A' é um n -tuplo cujas componentes podem ser qualquer elemento de S excepto (6, 6). Deste modo existem 35 valores possíveis para cada componente e portanto $(35)^n$ n -tuplos ao todo em A' . Visto que cada elemento de A' tem probabilidade $(\frac{1}{36})^n$, a soma de todas as probabilidades em A' é $(\frac{35}{36})^n$. Assim temos

$$P(A) = 1 - P(A') = 1 - \left(\frac{35}{36}\right)^n.$$

Para responder à questão de de Méré temos que verificar se $P(A)$ é maior ou menor que $\frac{1}{2}$ quando $n = 24$. A desigualdade $P(A) \geq \frac{1}{2}$ é equivalente a $1 - \left(\frac{35}{36}\right)^n \geq \frac{1}{2}$ ou $\left(\frac{35}{36}\right)^n \leq \frac{1}{2}$. Tomando logaritmos obtemos

$$n \log 35 - n \log 36 \leq -\log 2, \quad \text{ou} \quad n \geq \frac{\log 2}{\log 36 - \log 35} = 24.6+.$$

Por conseguinte $P(A) < \frac{1}{2}$ quando $n = 24$ e $P(A) > \frac{1}{2}$ quando $n \geq 25$. Não é, pois, vantajosa uma aposta de determinada importância a favor do aparecimento de um duplo seis, pelo menos uma vez, em 24 lançamentos de dois dados, em relação à aposta de uma mesma importância contrária a esse aparecimento.

A discussão que acabámos de efectuar sugere um método geral para tratar problemas de experiências aleatórias repetidas. Se numa dada experiência aleatória se repete duas ou mais vezes, o resultado pode considerar-se uma experiência composta. Mais geralmente, uma experiência composta pode considerar-se o resultado da realização sucessiva de duas ou mais experiências aleatórias distintas. Cada uma das experiências simples pode estar relacionada com cada uma das restantes ou podem ser aleatoriamente independentes, no sentido de que a probabilidade do resultado de qualquer uma delas não depende dos resultados das outras. Por uma questão de simplicidade, discutiremos como se podem combinar duas experiências independentes numa experiência composta. A generalização a mais do que duas experiências será evidente.

Para associar um espaço de probabilidade *bona fide* com uma experiência composta, devemos explicar como definir o novo conjunto fundamental S , a correspondente álgebra de Boole \mathcal{B} de subconjuntos de S , e a medida de probabilidade P definida em \mathcal{B} . Como no exemplo anterior recorreremos ao conceito de produto cartesiano.

Sejam $(S_1, \mathcal{B}_1, P_1)$, $(S_2, \mathcal{B}_2, P_2)$ dois espaços de probabilidade associados a duas experiências aleatórias E_1 e E_2 . Representemos por E a experiência composta para a qual o espaço amostra é $S_1 \times S_2$. Um resultado de E é um par (x, y) de S , com a primeira componente x um resultado de E_1 e a segunda componente y um resultado de E_2 . Se S_1 tem n elementos e S_2 tem m , então $S_1 \times S_2$ tem nm elementos.

Como nova álgebra de Boole \mathcal{B} tomamos a colecção de todos os subconjuntos de S . A seguir definimos a função de probabilidade P . Uma vez que S é finito podemos

definir $P(x, y)$ para cada ponto (x, y) de S e utilizar depois a aditividade para definir P para os subconjuntos de S . As probabilidades $P(x, y)$ podem definir-se de várias maneiras. Todavia, se as duas experiências E_1 e E_2 são aleatoriamente independentes, definimos P pela fórmula

$$(13.14) \quad P(x, y) = P_1(x)P_2(y) \quad \text{para todo } (x, y) \text{ em } S.$$

A justificação para esta definição é a seguinte. Consideremos dois acontecimentos particulares A e B do novo espaço S ,

$$A = \{(x_1, y_1), (x_1, y_2), \dots, (x_1, y_m)\}$$

e

$$B = \{(x_1, y_1), (x_2, y_1), \dots, (x_n, y_1)\}.$$

isto é, A é o conjunto de todos os pares de $S_1 \times S_2$ nos quais o primeiro elemento é x_1 , e B é o conjunto de todos os pares cujos segundo elemento é y_1 . A intersecção dos dois conjuntos A e B é o conjunto $\{(x_1, y_1)\}$. Se pensamos que o primeiro resultado, x_1 , não deve ter influência no segundo, y_1 , parece razoável exigir que A e B sejam independentes. Isto significa que teremos que definir a nova função de probabilidade P de tal modo que

$$(13.15) \quad P(A \cap B) = P(A)P(B).$$

Se decidirmos da forma como devemos atribuir as probabilidades $P(A)$ e $P(B)$, a equação (13.15) dir-nos-á como definir a probabilidade $P(A \cap B)$, ou seja a probabilidade $P(x_1, y_1)$. O acontecimento A ocorre se e só se o resultado da primeira experiência é x_1 . Uma vez que $P_1(x_1)$ é a sua probabilidade, parece natural atribuir o valor $P_1(x_1)$ também a $P(A)$. Analogamente, atribuímos a $P(B)$ o valor $P_2(y_1)$. A fórmula (13.15) dá-nos então

$$P(x_1, y_1) = P_1(x_1)P_2(y_1).$$

Tudo isto é, naturalmente, uma mera justificação para a atribuição de probabilidades em (13.14). O único modo de decidir se sim ou não (13.14) é uma distribuição aceitável de probabilidades pontuais é analisar se são verificadas as propriedades fundamentais das medidas de probabilidade. Cada número $P(x, y)$ é não negativo, e a soma de todas as probabilidades pontuais é igual a 1, uma vez que se tem

$$\sum_{(x, y) \in S} P(x, y) = \sum_{x \in S_1} P_1(x) \cdot \sum_{y \in S_2} P_2(y) = 1 \cdot 1 = 1.$$

Quando dizemos que numa experiência composta E é determinada por duas experiências aleatoriamente independentes E_1 e E_2 , queremos significar que o espaço de probabilidade (S, \mathcal{B}, P) é definido da maneira que acabámos de referir, ficando tal

"independência" reflectida no facto de que $P(x, y)$ é igual ao produto $P_1(x) \cdot P_2(y)$. Pode demonstrar-se que a atribuição de probabilidades em (13.14) implica a fórmula

$$(13.16) \quad P(U \times V) = P_1(U)P_2(V)$$

para todo o par de subconjuntos U de \mathcal{B}_1 e V de \mathcal{B}_2 . (Para um esboço de demonstração ver o Exercício 12 da Secção 12.23). Vamos deduzir algumas consequências importantes desta fórmula.

Seja A um acontecimento (na experiência composta E) da forma

$$A = C_1 \times S_2,$$

com $C_1 \in \mathcal{B}_1$. Cada resultado A é um par ordenado (x, y) em que x é restringido a ser um resultado de C_1 (na primeira experiência E_1), mas y pode ser qualquer resultado de S_2 (na segunda experiência E_2). Se aplicarmos (13.16) encontramos:

$$P(A) = P(C_1 \times S_2) = P_1(C_1)P_2(S_2) = P_1(C_1),$$

uma vez que $P_2(S_2) = 1$. Deste modo a definição de P atribui a A a mesma probabilidade que P_1 atribui a C_1 . Por este motivo, um tal acontecimento A diz-se ser *determinado pela primeira experiência E_1* . De modo análogo, se B é um acontecimento de E da forma

$$B = S_1 \times C_2,$$

com $C_2 \in \mathcal{B}_2$, temos

$$P(B) = P(S_1 \times C_2) = P_1(S_1)P_2(C_2) = P_2(C_2)$$

e B diz-se ser *determinado pela segunda experiência E_2* . Vamos agora provar, utilizando (13.16), que dois tais acontecimentos A e B são *independentes*, isto é, que se tem

$$(13.17) \quad P(A \cap B) = P(A)P(B).$$

Observe-se primeiramente que

$$\begin{aligned} A \cap B &= \{(x, y) \mid (x, y) \in C_1 \times S_2 \text{ e } (x, y) \in S_1 \times C_2\} \\ &= \{(x, y) \mid x \in C_1 \text{ e } y \in C_2\} \\ &= C_1 \times C_2. \end{aligned}$$

Logo, devido a (13.16), temos

$$(13.18) \quad P(A \cap B) = P(C_1 \times C_2) = P_1(C_1)P_2(C_2).$$

Já que $P_1(C_1) = P(A)$ e $P_2(C_2) = P(B)$, obtemos (13.17). Observe-se que a equação (13.18) mostra também que podemos calcular a probabilidade $P(A \cap B)$ como um produto das probabilidades em cada um dos espaços amostras S_1 e S_2 ; logo não são necessários quaisquer cálculos com probabilidades na experiência composta.

A generalização a experiências compostas, definidas por n experiências E_1, E_2, \dots, E_n procede-se do mesmo modo. Os pontos neste novo espaço amostra são n -tuplos (x_1, x_2, \dots, x_n) e as probabilidades pontuais definem-se como produto das probabilidades dos resultados particulares

$$(13.19) \quad P(x_1, x_2, \dots, x_n) = P_1(x_1)P_2(x_2) \cdots P_n(x_n).$$

Quando se adopta esta definição de P dizemos que E é determinada por n experiências aleatórias independentes E_1, E_2, \dots, E_n . No caso particular em que todas as experiências estão associadas com o mesmo espaço de probabilidade, a experiência composta E diz-se constituir um exemplo de *provas independentes repetidas sob idênticas condições*. Na Secção seguinte analisamos um tal exemplo.

13.16 Esquema de Bernoulli

Um exemplo importante de uma experiência composta foi exaustivamente estudado por Jacob Bernoulli e é actualmente conhecido por *esquema de Bernoulli*. Este consiste numa sucessão de experiências idênticas, realizadas nas mesmas condições, sendo cada resultado aleatoriamente independente dos restantes. Em cada experiência são possíveis dois tipos de resultado, correntemente designados por *acontecimento favorável* e *acontecimento contrário*; a probabilidade do acontecimento favorável representa-se por p e a do acontecimento contrário por q . Evidentemente que $q = 1 - p$. O principal resultado associado com o esquema de Bernoulli é o seguinte:

TEOREMA 13.3. FÓRMULA DE BERNOULLI. *A probabilidade de ocorrência do acontecimento favorável k vezes em n experiências realizadas num esquema de-Bernoulli é*

$$(13.20) \quad \binom{n}{k} p^k q^{n-k},$$

com $\binom{n}{k}$ o coeficiente binomial, $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

Demonstração. Representemos o acontecimento favorável por S e o acontecimento contrário por F e consideremos uma sucessão particular de n resultados. Isto pode representar-se por um n -tuplo

$$(x_1, x_2, \dots, x_n),$$

onde cada x_i é ou um S ou F . O acontecimento A no qual estamos interessados é uma colecção de todos os n -tuplos que contêm precisamente k vezes o S e $n - k$ vezes o F . Calculemos a probabilidade pontual de um n -tuplo particular de A . A probabilidade de

cada S é p e a de cada F é q . Por conseguinte, por (13.19), a probabilidade de cada n -tuplo particular de A é o produto de k factores iguais a p por $n - k$ factores iguais a q , isto é,

$$P(x_1, x_2, \dots, x_n) = p^k q^{n-k} \quad \text{se } (x_1, x_2, \dots, x_n) \in A.$$

Deste modo, para calcularmos $P(A)$ necessitamos unicamente determinar o número de elementos de A e multiplicar este número por $p^k q^{n-k}$. Mas o número de elementos de A é muito simplesmente o número de maneiras distintas de colocar k vezes o S em todas as n possíveis posições do n -tuplo. Este número é o mesmo que dos subconjuntos de k elementos que podem formar-se a partir do conjunto de n elementos; sabemos já que tal número é dado por $\binom{n}{k}$. Deste modo, se adicionamos as probabilidades pontuais para todos os elementos de A obtemos

$$P(A) = \binom{n}{k} p^k q^{n-k}.$$

EXEMPLO 1. Lança-se ao ar 50 vezes uma moeda imparcial. Calcular a probabilidade de obter precisamente 25 vezes a face cara.

Resolução. Interpretamos este problema como um esquema de Bernoulli no qual se realizam n experiências e onde o acontecimento favorável é a obtenção da face “cara”. Uma vez que a moeda é “perfeita” atribuímos as probabilidades $p = q = \frac{1}{2}$ e a fórmula (13.20) dá-nos $\binom{50}{k} \left(\frac{1}{2}\right)^{50}$ para a probabilidade de obter exactamente k caras em 50 lançamentos. Em particular, quando $k = 25$ obtemos

$$\binom{50}{25} \left(\frac{1}{2}\right)^{50} = \frac{50!}{25! 25!} \left(\frac{1}{2}\right)^{50}.$$

Para exprimir este número na forma decimal é preferível usar logaritmos, visto que são correntes as tábuas de logaritmos de factoriais. Se representamos o número em questão por P , uma tábua de logaritmos decimais dá-nos

$$\begin{aligned} \log P &= \log 50! - 2 \log 25! - 50 \log 2 \\ &= 64.483 - 50.381 - 15.052 = -0.950 = 0.05 - 1.00 \\ &= \log 1.12 - \log 10 = \log 0.112, \end{aligned}$$

pelo que $P = 0.112$.

EXEMPLO 2. Qual a probabilidade de obter pelo menos r vezes o acontecimento favorável em n experiências num esquema de Bernoulli?

Resolução. Representemos por A_k o acontecimento “precisamente k vezes o acontecimento favorável em n experiências”. Então o acontecimento E no qual estamos interessados é a união

$$E = A_r \cup A_{r+1} \cup \dots \cup A_n.$$

Uma vez que A_k são disjuntos, encontramos

$$P(E) = \sum_{k=r}^n P(A_k) = \sum_{k=r}^n \binom{n}{k} p^k q^{n-k}.$$

Porque

$$\sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = (p + q)^n = 1,$$

a probabilidade do acontecimento contrário E' pode calcular-se por:

$$P(E') = 1 - P(E) = \sum_{k=0}^{r-1} \binom{n}{k} p^k q^{n-k}.$$

A última soma dá-nos a probabilidade da ocorrência de, quando muito, $r - 1$ acontecimentos favoráveis em n experiências.

13.17. O número mais favorável de ocorrências do acontecimento favorável em n experiências dum esquema de Bernoulli

Lança-se um par de dados 28 vezes. Qual é o número mais provável de vezes em que se obterá uma soma de pontos igual a sete? Para resolver o problema designemos por $f(k)$ a probabilidade de obter precisamente k vezes a soma de pontos 7 em 28 lançamentos. A probabilidade de obter um sete num lançamento é $\frac{1}{6}$. A fórmula de Bernoulli diz-nos que

$$f(k) = \binom{28}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{28-k}.$$

Queremos determinar qual o valor (ou valores) de k , na sucessão $k = 0, 1, 2, \dots, 28$ que tornam $f(k)$ máximo. O teorema que se segue dá uma resposta a esta questão para qualquer sucessão de experiências num esquema de Bernoulli.

TEOREMA 13.4. *Dados um inteiro $n \geq 1$ e um real p , $0 < p < 1$, considerem-se o conjunto de números*

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \text{para } k = 0, 1, \dots, n.$$

(a) Se $(n+1)p$ não é inteiro, o valor máximo de $f(k)$ ocorre para, precisamente, um valor de k :

$$k = [(n+1)p], \quad \text{o maior inteiro} < (n+1)p.$$

(b) Se $(n+1)p$ é inteiro, o valor máximo de $f(k)$ tem lugar para, precisamente, dois valores de k :

$$k = (n+1)p \quad \text{e} \quad k = (n+1)p - 1.$$

Demonstração. Para estudarmos o comportamento de $f(k)$ consideremos o cociente

$$r(k) = \frac{f(k)}{f(k+1)} = \frac{k+1}{n-k} \frac{1-p}{p}$$

para $k=0, 1, \dots, n-1$. A função $r(k)$ é estritamente crescente pelo que se tem

$$0 < r(0) < r(1) < \dots < r(n-1).$$

Consideremos seis casos, representados na figura 13.2: Nos três primeiros casos mostra-se que $f(k)$ toma o valor máximo precisamente para um só valor de k . Nos restantes casos $f(k)$ toma o seu valor máximo para dois valores consecutivos de k .

CASO 1. $r(0) > 1$. Neste caso $r(k) > 1$ para todo o k pelo que se tem

$$f(0) > f(1) > \dots > f(n).$$

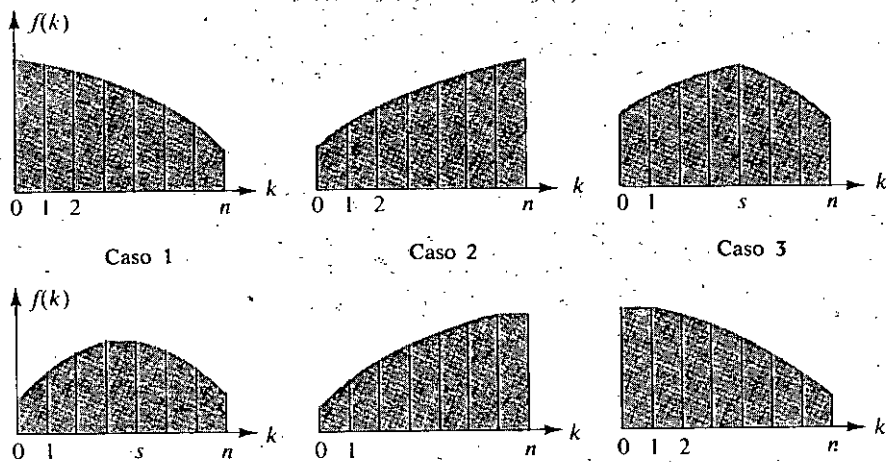


FIG. 13.2. Cálculo do número mais provável de ocorrências do acontecimento favorável em n experiências num esquema de Bernoulli.

Deste modo o maior valor de $f(k)$ ocorre unicamente para $k = 0$. Também, $r(0) = (1-p)/(np) > 1$, pelo que $1-p > np$, $(n+1)p < 1$, logo $[(n+1)p] = 0$.

CASO 2. $r(n-1) < 1$. Neste caso $r(k) < 1$ para todo k pelo que $f(0) < f(1) < \dots < f(n)$ e o máximo valor de $f(k)$ ocorre unicamente para $k = n$. Uma vez que $r(n-1) = n(1-p)/p < 1$, temos $n - np < p$, logo $n < (n+1)p < n+1$, onde $[(n+1)p] = n$.

CASO 3. $r(0) < 1$, $r(n-1) > 1$, e $r(k) \neq 1$ para todo k . Neste caso existe um único inteiro s , $0 < s < n$, tal que $r(s-1) < 1$ e $r(s) > 1$. A função $f(k)$ cresce no intervalo $0 \leq k \leq s$ e decrescente no intervalo $s \leq k \leq n$. Portanto $f(k)$ tem um único máximo em $k = s$. Visto que $r(s-1) = s(1-p)/(np-sp+p) < 1$ temos $s < (n+1)p$. A desigualdade $r(s) > 1$ mostra que $(n+1)p < s+1$, logo $[(n+1)p] = s$.

Observe-se que em cada um dos três primeiros casos o valor máximo de $f(k)$ ocorre quando $k = [(n+1)p]$; também $(n+1)p$ não é inteiro em qualquer destes casos.

CASO 4. $r(0) < 1$, $r(n-1) > 1$, e $r(s-1) = 1$ para algum s , $2 \leq s < n$. Neste caso $f(k)$ cresce para $0 \leq k \leq s-1$ e decresce para $s \leq k \leq n$. O máximo de $f(k)$ ocorre duas vezes, quando $k = s-1$ e $k = s$. A equação $r(s-1) = 1$ implica $(n+1)p = s$.

CASO 5. $r(n-1) = 1$. Neste caso $r(k) < 1$ para $k \leq n-2$, pelo que $f(k)$ cresce no intervalo $0 \leq k \leq n-1$, e $f(n-1) = f(n)$. Logo o máximo de $f(k)$ ocorre duas vezes, quando $k = n-1$ e quando $k = n$. A igualdade $r(n-1) = 1$ implica que $(n+1)p = n$.

CASO 6. $f(0) = 1$. Neste caso $r(k) > 1$ para $k \geq 1$, pelo que $f(k)$ decresce no intervalo $1 \leq k \leq n$. O valor máximo de $f(k)$ ocorre duas vezes, quando $k = 0$ e $k = 1$. A igualdade $r(0) = 1$ implica que $(n+1)p = 1$.

Em cada um dos três últimos casos o valor máximo de $f(k)$ ocorre para $k = (n+1)p$ e $k = (n+1)p - 1$. Está, pois, completada a demonstração.

EXEMPLO 1. Lança-se um par de dados 28 vezes. Qual é o número de ocorrências mais provável da obtenção de uma soma de pontos igual a 7?

Resolução. Aplicamos o teorema 13.4 com $n = 28$, $p = 1/6$ e $(n+1)p = 29/6$. Este valor é não inteiro, pelo que o maior valor de $f(k)$ ocorre quando $k = [29/6] = 4$.

Observação. Se os dados forem lançados 29 vezes existem duas soluções $k = 4$ e $k = 5$.

EXEMPLO 2. Determinar o menor n tal que, lançando um par de dados n vezes, a probabilidade de obter precisamente 4 vezes um total de pontos igual a sete seja pelo menos tão grande como a probabilidade de obter qualquer outro número de setes.

Resolução. Fizemos $p = \frac{1}{6}$ no teorema 13.4. Pretendemos que o máximo valor de $f(k)$ ocorra quando $k = 4$. Isto exige ou que $[(n+1)p] = 4$, $(n+1)p = 4$, ou que $(n+1)p - 1 = 4$. O menor valor de n satisfazendo a qualquer destas relações é $n = 23$.

13.18. Exercícios

1. Lança-se uma moeda duas vezes ao ar, sendo p_1 a probabilidade de saída da face cara (H) no primeiro lançamento e p_2 no segundo. Consideremos esta como uma experiência composta determinada por duas experiências aleatoriamente independentes, e seja o conjunto fundamental

$$S = \{(H, H), (H, T), (T, H), (T, T)\}.$$

- (a) Calcular a probabilidade de cada elemento de S .
 (b) Podem atribuir-se as probabilidades p_1 e p_2 de modo que

$$P(H, H) = \frac{1}{9}, \quad P(H, T) = P(T, H) = \frac{2}{9}, \quad P(T, T) = \frac{4}{9}?$$

- (c) Podem atribuir-se as probabilidades p_1 e p_2 de modo que

$$P(H, H) = P(T, T) = \frac{1}{3}, \quad P(H, T) = P(T, H) = \frac{1}{6}?$$

- (d) Considerar os quatro acontecimentos seguintes (subconjuntos de S):

$$\begin{aligned} H_1: & \text{"cara" no primeiro lançamento,} \\ H_2: & \text{"cara" no segunda lançamento,} \\ T_1: & \text{"cruz" no primeiro lançamento,} \\ T_2: & \text{"cruz" no segundo lançamento.} \end{aligned}$$

Determinar destes quatro acontecimentos quais os pares que são independentes.

Em cada um dos Exercícios 2 a 12 definir o conjunto fundamental, a atribuição de probabilidades e o acontecimento cuja probabilidade se calcula.

2. Um estudante tem que responder a um teste de 10 questões com as respostas verdadeiro e falso. Como não está preparado para dar resposta às perguntas resolve responder-lhes ao acaso. Para isso lança uma moeda ao ar e serve-se do resultado para decidir da sua resposta.
- (a) Calcular a probabilidade de que responda correctamente pelo menos cinco vezes.
 (b) Calcular a probabilidade de que responda correctamente pelo menos nove vezes.
 (c) Qual será o menor n tal que a probabilidade de acertar pelo menos n respostas seja menor que $\frac{1}{2}$?
3. Lançam-se simultaneamente dez dados perfeitos. Qual a probabilidade de obter precisamente três senas.
4. Lança-se ao ar cinco vezes uma moeda imparcial. Qual é a probabilidade de obter (a) três vezes a face "cara"? (b) pelo menos três vezes a face "cara"? (c) quando muito uma vez face "cara"?
5. Um homem afirmava ter uma vara com a qual localizava jazigos de petróleo. O laboratório de Geologia de Caltech realiza a seguinte experiência para pôr à prova a sua afirmação. Colocam-no numa sala onde havia 10 barris fechados. Dizem-lhe que cinco deles contêm petróleo e outros cinco contêm água. O seu trabalho consiste em identificar os que contêm petróleo e os que contêm água.
- (a) Qual a probabilidade de que, apenas devido ao acaso, ele localize correctamente os cinco barris de petróleo?
 (b) Qual a probabilidade de que, apenas devido ao acaso, localize unicamente três dos barris de petróleo.

6. Uma velha senhora de Pasadena afirma que provando uma chávena de chá com leite pode dizer qual foi o que se colocou primeiro na chávena: se o leite, se o chá. Tal afirmação é posta à prova obrigando-a a provar e classificar 10 pares de chávenas de chá, cada par contendo uma chávena com a mistura feita dos dois modos possíveis. Seja p a sua probabilidade "verdadeira" de classificar um par de chávenas correctamente. (Se ela é hábil, p é substancialmente maior que $\frac{1}{2}$; se o não é, $p \leq \frac{1}{2}$). Supõe-se que os 10 pares de taças são classificadas sob condições idênticas e independentes.
 - (a) Calcular, em função de p , a probabilidade de que ela classifique correctamente pelo menos oito dos dez pares de taças.
 - (b) Calcular esta probabilidade quando $p = \frac{1}{2}$.
7. (Outro problema do Cavaleiro de Méré). Determinar se é ou não vantajoso jogar, apostando iguais importâncias, a que saia pelo menos uma vez a face 6 em quatro lançamentos de um dado imparcial. [Sugestão: Mostrar que a probabilidade de obter pelo menos uma vez a face 6 em n lançamentos é $1 - (\frac{5}{6})^n$].
8. Uma urna contém b bolas brancas e p bolas pretas. Se $k \leq n$, calcular a probabilidade de extrair k bolas brancas em n tiragens, com reposição, de cada vez, da bola extraída.
9. Lançam-se dois dados simultaneamente oito vezes. Calcular a probabilidade de obter precisamente três vezes uma soma de pontos igual a 11.
10. Lança-se uma moeda 10 vezes ou 10 moedas simultaneamente uma vez e conta-se o número de vezes que se obteve "caras". Determinar a probabilidade de que tal número seja pelo menos seis.
11. Depois de uma longa série de testes referentes a um dado tipo de foguetão concluiu-se que em cerca de 5 % dos lançamentos se produzia um mau funcionamento que seria causa do fracasso do lançamento. Calcular a probabilidade de, em 10 lançamentos, existir pelo menos um insucesso.
12. Lança-se uma moeda repetidamente ao ar. Calcular a probabilidade de que o número total de "caras" seja pelo menos seis antes que o número total de "cruzes" seja 5.
13. O Exercício 12 pode generalizar-se do modo seguinte: mostrar que a probabilidade de que ocorram pelo menos m acontecimentos favoráveis, antes que ocorram n acontecimentos contrários, numa sucessão de experiências de Bernoulli, é

$$\sum_{k=m}^{m+n-1} \binom{m+n-1}{k} p^k q^{m+n-k-1}.$$

14. Determinar todos os valores de n com a seguinte propriedade: Se um par de dados perfeitos é lançado n vezes, a probabilidade de obter exactamente 10 setes é, pelo menos, tão grande como a probabilidade de obter qualquer outro número de setes.
15. Uma máquina binária de moedas tem três rodas idênticas e independentes. Quando se joga com a máquina os resultados possíveis são ternos ordenados (x, y, z) em que cada letra x, y , ou z pode ser 0 ou 1. Em cada roda a probabilidade de 0 é p e a probabilidade de 1 é $1 - p$, com $0 < p < 1$. A máquina paga 60\$ 00 se o resultado é $(1, 1, 1)$ ou $(0, 0, 0)$; paga 30\$ 00 se o resultado for $(1, 1, 0)$; nos restantes casos não paga nada. Seja $f(p)$ a probabilidade de que a máquina pague 30\$ 00 ou mais quando se joga uma vez.
 - (a) Calcular $f(p)$
 - (b) Definir o "pago total" pela soma $\sum_{x \in S} g(x)P(x)$, em que S é o conjunto fundamental, $P(x)$ é a probabilidade do resultado x e $g(x)$ o número de vezes 30\$ 00 pagos pelo resultado x . Calcular o valor de p para o qual o "pago total" seja mínimo.

13.19. Conjuntos numeráveis e não numeráveis

Até aqui considerámos unicamente a teoria das probabilidades para conjuntos fundamentais finitos. Pretendemos agora generalizar a teoria para conjuntos infinitos. Com estas finalidade é necessário distinguir entre dois tipos de conjuntos infinitos, os *numeráveis* e os *não numeráveis*. Vamos passar a analisar estes dois conceitos.

Para contar os elementos de um conjunto finito põe-se em correspondência o conjunto, elemento a elemento, com o conjunto dos números naturais $\{1, 2, \dots, n\}$. A comparação dos “tamanhos” de dois conjuntos mediante a correspondência entre eles, elemento a elemento, substitui a contagem dos elementos quando se trata de conjuntos infinitos. A este processo pode dar-se uma clara formulação matemática empregando o conceito de função:

DEFINIÇÃO: *Dois conjuntos A e B dizem-se em correspondência biunívoca se existir uma função f com as seguintes propriedades:*

(a) *O domínio de f é A e o contradomínio é B .*

(b) *Se x e y são elementos distintos de A , então $f(x)$ e $f(y)$ são elementos distintos de B , isto é, para todo x e y de A ,*

$$(13.21) \quad x \neq y \quad \text{implica} \quad f(x) \neq f(y).$$

Uma função que satisfaça à propriedade (13.21) diz-se *biunívoca* sobre A . Dois conjuntos A e B em correspondência biunívoca dizem-se também *equivalentes*, e exprimimos esta propriedade pelo símbolo $A \sim B$. É evidente que todo o conjunto A é equivalente a si próprio, uma vez que podemos fazer $f(x) = x$ para todo x em A .

Um conjunto pode ser equivalente a um subconjunto próprio dele mesmo. Por exemplo, o conjunto $P = \{1, 2, 3, \dots\}$ formado pelos inteiros positivos, é equivalente ao subconjunto próprio $Q = \{2, 4, 6, \dots\}$ formado pelos inteiros pares. Neste caso a função biunívoca que os faz equivalentes é dada por $f(x) = 2x$ para todo x de P .

Se $A \sim B$ é fácil de provar que $B \sim A$. Com efeito, se f é biunívoca sobre A e se o contra-domínio de f é B , então para cada b em B existe precisamente um a em A tal que $f(a) = b$. Deste modo podemos definir uma função inversa g em B do modo seguinte: Se $b \in B$, $g(b) = a$, onde a é o único elemento de A tal que $f(a) = b$. Esta função g é biunívoca sobre B , e o seu contradomínio é A ; logo $B \sim A$. Esta propriedade da equivalência é conhecida por *simetria*.

$$(13.22) \quad A \sim B \quad \text{implica} \quad B \sim A.$$

É igualmente fácil de demonstrar que a equivalência goza da seguinte propriedade, conhecida por *propriedade transitiva*:

$$(13.23) \quad A \sim B \quad \text{e} \quad B \sim C \quad \text{implica} \quad A \sim C.$$

No Exercício 2 da Secção 13.20 pède-se a demonstração desta propriedade.

Um conjunto S chama-se *finito*, e diz-se conter n elementos, se

$$S \sim \{1, 2, \dots, n\}.$$

O conjunto vazio considera-se também finito. Conjuntos que não são finitos dizem-se *infinitos*. Um conjunto S chama-se *infinito numerável* se é equivalente ao conjunto dos números naturais, isto é, se

$$(13.24) \quad S \sim \{1, 2, 3, \dots\}.$$

Neste caso existe uma função f a qual estabelece uma correspondência biunívoca entre os inteiros positivos e os elementos de S , por isso o conjunto S pode exprimir-se na forma

$$S = \{f(1), f(2), f(3), \dots\}.$$

Recorremos muitas vezes aos subíndices e representamos $f(k)$ por a_k (ou por outra notação semelhante) e escrevemos $S = \{a_1, a_2, a_3, \dots\}$. O importante aqui é que a correspondência (13.24) permite-nos usar os números naturais como “marcas” dos elementos de S .

Um conjunto diz-se *numerável*, no sentido lato, se fôr finito ou infinito numerável. Um conjunto que não seja numerável diz-se *não numerável*. Muitas operações com conjuntos numeráveis produzem conjuntos numeráveis. Por exemplo, temos as seguintes propriedades:

- (a) Cada subconjunto de um conjunto numerável é numerável.
- (b) A intersecção de qualquer família de conjuntos numeráveis é numerável.
- (c) A união de uma família numerável de conjuntos numeráveis é numerável.
- (d) O produto cartesiano de um número finito de conjuntos numeráveis é numerável.

Uma vez que pouco utilizaremos, neste livro, os conjuntos infinitos numeráveis, não apresentamos a demonstração detalhada destas propriedades†. Em seu lugar apresentaremos um certo número de exemplos para mostrarmos como estas propriedades podem utilizar-se para construir novos conjuntos numeráveis a partir de conjuntos dados.

EXEMPLO 1. O conjunto S de todos os inteiros (positivos, negativos e zero) é numerável.

† As demonstrações estão delineadas nos Exercícios 3 a 8 da Secção 13.20.

Demonstração. Se $n \in S$, $f(n) = 2n$ sendo n positivo, e $f(n) = 2n + 1$ sendo n negativo ou zero. O domínio de f é S e o seu contradomínio é o conjunto dos inteiros positivos. Visto que f é biunívoco em S , consequentemente S é numerável.

EXEMPLO 2. O conjunto R de todos os números racionais é numerável.

Demonstração. Para todo o inteiro $n \geq 1$ fixo, seja S_n o conjunto de números racionais da forma x/n , onde x pertence ao conjunto S do exemplo 1. Cada S_n é equivalente a S [fazer $f(t) = nt$ se $t \in S_n$], consequentemente cada S_n é numerável. Visto que R é a união de todos os S_n pela propriedade (c), R é numerável.

Nota. Se $\mathcal{F} = \{A_1, A_2, A_3, \dots\}$ é uma família numerável de conjuntos, e a união deles se exprime com o símbolo.

$$\bigcup_{k=1}^{\infty} A_k \quad \text{ou} \quad A_1 \cup A_2 \cup A_3 \cup \dots$$

EXEMPLO 3. Seja A um conjunto infinito numerável, por exemplo $A = \{a_1, a_2, a_3, \dots\}$. Para todo o inteiro $n \geq 1$, seja \mathcal{F}_n a família dos subconjuntos de A com n elementos, isto é, seja

$$\mathcal{F}_n = \{S \mid S \subseteq A \text{ e } S \text{ tem } n \text{ elementos}\}.$$

Então cada \mathcal{F} é numerável.

Demonstração. Se S é um subconjunto de A com n elementos, podemos escrever

$$S = \{a_{k_1}, a_{k_2}, \dots, a_{k_n}\},$$

com $k_1 < k_2 < \dots < k_n$. Seja $f(S) = (a_{k_1}, a_{k_2}, \dots, a_{k_n})$, isto é, seja f uma função que associa a S o n -tuplo ordenado $(a_{k_1}, a_{k_2}, \dots, a_{k_n})$. O domínio de f é \mathcal{F}_n e o seu contradomínio, que representamos por T_n , é um subconjunto do produto cartesiano $C_n = A \times A \times \dots \times A$ (n vezes). Já que A é numerável, o mesmo se verifica com C_n [propriedade (d)] e por isso T_n é-o também [propriedade (a)]. Mas $T_n \sim \mathcal{F}_n$ porque f é biunívoca. Isto prova que \mathcal{F}_n é numerável.

EXEMPLO 4. A família de todos os subconjuntos finitos de um conjunto numerável é numerável.

Demonstração. O resultado é evidente se o subconjunto dado é finito. Suponhamos, então, que o conjunto dado (o conjunto A) é infinito numerável, e seja \mathcal{F} a família de todos os subconjuntos finitos de A .

$$\mathcal{F} = \{S \mid S \subseteq A \text{ e } S \text{ é finito}\}.$$

Então \mathcal{F} é a união de todas as famílias \mathcal{F}_n do Exemplo 3; logo, pela propriedade (c), \mathcal{F} é numerável.

EXEMPLO 5. A família de *todos* os subconjuntos de um conjunto infinito numerável é não numerável.

Demonstração. Seja A o conjunto numerável dado e \mathcal{A} a família de todos os subconjuntos de A . Suponhamos que \mathcal{A} é numerável e chegaremos a um absurdo. Se \mathcal{A} é numerável, então $\mathcal{A} \sim A$ e por conseguinte existe uma função biunívoca f cujo domínio é A e o contradomínio é \mathcal{A} . Assim para cada a em A , o valor da função $f(a)$ é um subconjunto de A . Este subconjunto pode ou não conter o elemento a . Representamos por B o conjunto de elementos a tais que $a \notin f(a)$. Assim,

$$B = \{a \mid a \in A \text{ e } a \notin f(a)\}.$$

Se B é um subconjunto de A , deve pertencer à família \mathcal{A} . Isto significa que $B = f(b)$ para algum b de A . Existem agora duas possibilidades: (i) $b \in B$, ou (ii) $b \notin B$. Se $b \in B$, então pela definição de B temos $b \notin f(b)$, o que está em contradição visto que $f(b) = B$. Portanto (i) é impossível. No caso (ii), $b \notin B$, o que significa $b \in f(b)$. Isto contradiz a definição de B , pelo que a hipótese (ii) é também impossível. Deste modo a afirmação de que A é numerável conduz-nos a uma contradição e portanto temos que concluir que \mathcal{A} é não numerável.

Apresentamos a seguir um exemplo de conjunto não numerável que é mais simples do que o Exemplo 5.

EXEMPLO 6. O conjunto dos números reais x tais que $0 < x < 1$ é não numerável.

Demonstração. Mais uma vez, suponhamos que o conjunto é numerável e chegamos a uma contradição. Se o conjunto é numerável podemos dispor os seus elementos do modo seguinte: $\{x_1, x_2, \dots\}$. Construiremos agora um número real y satisfazendo a $0 < y < 1$ e que não estará nesta lista. Com esse objectivo escrevemos cada elemento x_n na forma decimal

$$x_n = 0.a_{n,1}a_{n,2}a_{n,3}\dots,$$

onde cada $a_{n,i}$ é um dos inteiros do conjunto $\{0, 1, 2, \dots, 9\}$. Seja y o número real cujo desenvolvimento decimal é:

$$y = 0.y_1y_2y_3\dots,$$

onde

$$y_n = \begin{cases} 1 & \text{se } a_{n,n} \neq 1, \\ 2 & \text{se } a_{n,n} = 1. \end{cases}$$

Então nenhum elemento do conjunto $\{x_1, x_2, x_3, \dots\}$ pode ser igual a y , porque y difere de x_1 na primeira casa decimal, difere de x_2 na segunda e, em geral, difere de x_k na k -ésima casa decimal. (Uma situação tal como $x_n = 0,249999\dots$ e $y = 0,25000\dots$ não pode verificar-se aqui devido ao modo como é escolhido y_n . Visto que y satisfaz a $0 < y < 1$, temos uma contradição e por conseguinte o conjunto dos números reais no intervalo aberto $(0,1)$ não é numerável.

13.20. Exercícios

- Seja $P = \{1, 2, 3, \dots\}$ o conjunto dos números naturais. Para cada um dos seguintes conjuntos, dar uma função biunívoca f cujo domínio seja P e o contradomínio o conjunto que se indica:
 - $A = \{2, 4, 6, \dots\}$, o conjunto dos inteiros positivos pares.
 - $B = \{3, 3^2, 3^3, \dots\}$, o conjunto das potências de 3.
 - $C = \{2, 3, 5, 7, 11, 13, \dots\}$, o conjunto dos números primos. [Nota: uma parte da demonstração consiste em provar que C é um conjunto infinito].
 - $P \times P$, o produto cartesiano de P por si próprio.
 - O conjunto dos inteiros da forma $2^m 3^n$ com m e n inteiros e positivos.
- Demonstrar a propriedade transitiva da equivalência de conjuntos

$$\text{Se } A \sim B \quad \text{e} \quad B \sim C, \quad \text{então} \quad A \sim C.$$

[Sugestão: Se f faz A equivalente a B e se g faz B equivalente a C , provar que a função composta $h = g \circ f$ faz A equivalente a C].

Os Exercícios 3 a 8 são destinados à demonstração das propriedades (a), (b), (c), (d) dos conjuntos numeráveis apresentados na Secção 13.19.

- Provar que todo o subconjunto de um conjunto numerável é numerável. [Sugestão: Suponhamos que S é um conjunto infinito numerável, por exemplo $S = \{x_1, x_2, x_3, \dots\}$, e seja A um subconjunto infinito de S . Seja $k(1)$ o menor inteiro positivo m tal que $x_m \in A$. Supostos definidos $k(1), k(2), \dots, k(n-1)$, seja $k(n)$ o menor inteiro positivo $m > k(n-1)$ tal que $x_m \in A$. Seja $f(n) = x_{k(n)}$. Mostrar que f é uma função biunívoca cujo domínio é o conjunto dos inteiros positivos e o contra domínio é A . Isto demonstra a proposição na hipótese de que S é infinito numerável. Construir outra demonstração para S finito].
- Provar que a intersecção de qualquer família de conjuntos numeráveis é numerável. [Sugestão: Usar o resultado do Exercício 3].
- Seja $P = \{1, 2, 3, \dots\}$ o conjunto dos números naturais:
 - Demonstrar que o produto cartesiano $P \times P$ é numerável. [Sugestão: Seja Q o conjunto dos inteiros e positivos da forma $2^m 3^n$, com n e m inteiros positivos. Então $Q \subseteq P$, pelo que Q é numerável (pelo Exercício 3). Se $(m, n) \in P \times P$, seja $f(m, n) = 2^m 3^n$ e usemos esta função para mostrar que $P \times P \sim Q$.]
 - Deduzir da alínea (a) que o produto cartesiano de dois conjuntos numeráveis é numerável. Usar depois o método de indução para generalizar o resultado a n conjuntos numeráveis.
- Seja $\mathcal{B} = \{B_1, B_2, B_3, \dots\}$ uma família numerável de conjuntos disjuntos ($B_i \cap B_j = \emptyset$ se $i \neq j$) tal que cada B_n é numerável. Demonstrar que a reunião $\bigcup_{k=1}^{\infty} B_k$ é também numerável. [Sugestão: Seja $B_n = \{b_{1,n}, b_{2,n}, b_{3,n}, \dots\}$ e $S = \bigcup_{k=1}^{\infty} B_k$. Se $x \in S$, então $x = b_{m,n}$ para algum

único par (m, n) e podemos definir $f(x) = (m, n)$. Utilizar esta f para demonstrar que S é equivalente a um subconjunto de $P \times P$ e deduzir (em virtude do Exercício 5) que S é numerável].

7. Seja $\mathcal{A} = \{A_1, A_2, A_3, A_4, \dots\}$ uma família numerável de conjuntos, e seja $\mathcal{B} = \{B_1, B_2, B_3, \dots\}$ definido do modo seguinte: $B_1 = A_1$ e, para $n > 1$,

$$B_n = A_n - \bigcup_{k=1}^{n-1} A_k.$$

Isto é, B_n consta dos elementos de A_n que não pertencem aos conjuntos precedentes A_1, \dots, A_{n-1} . Provar que \mathcal{B} é uma família de conjuntos disjuntos ($B_i \cap B_j = \emptyset$ quando $i \neq j$) e que

$$\bigcup_{k=1}^{\infty} A_k = \bigcup_{k=1}^{\infty} B_k.$$

Isto permite-nos exprimir a união de qualquer família numerável de conjuntos com a união de uma família numerável de conjuntos *disjuntos*.

8. Se \mathcal{F} é uma família numerável de conjuntos numeráveis, provar que a união de todos os conjuntos de \mathcal{F} é numerável. [Sugestão: Utilizar os Exercícios 6 e 7].
9. Provar que os seguintes conjuntos são numeráveis:
- (a) O conjunto de todos os intervalos de eixo real cujos extremos são números racionais.
 - (b) O conjunto de todos os círculos no plano com raios racionais e os centros com coordenadas racionais.
 - (c) Qualquer conjunto de intervalos disjuntos de comprimento positivo.
10. Provar que os conjuntos seguintes são não numeráveis:
- (a) O conjunto dos números irracionais do intervalo $(0, 1)$.
 - (b) O conjunto de todos os intervalos de comprimento positivo.
 - (c) O conjunto de todas as sucessões cujos termos são os inteiros 0 e 1. (Recorde-se que uma sucessão é uma função cujo domínio é o conjunto dos inteiros positivos).

13.21. Definição de probabilidade para conjuntos fundamentais infinitos numeráveis

Trata-se aqui de generalizar a definição de probabilidade para conjuntos fundamentais infinitos numeráveis. Seja S um tal conjunto infinito numerável e \mathcal{B} a álgebra de Boole de subconjuntos de S . Definimos uma medida de probabilidade P em \mathcal{B} como fizemos para o caso finito, excepto em que se exigirá a aditividade numerável além da aditividade finita. Quer dizer, para cada família infinita numerável $\{A_1, A_2, \dots\}$ de elementos de \mathcal{B} , exigimos que

$$(13.25) \quad P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k) \quad \text{se} \quad A_i \cap A_j = \emptyset \quad \text{sempre que} \quad i \neq j.$$

Funções de conjunto finitamente aditivas que verificam (13.25) dizem-se de *aditividade numerável* (ou completamente aditivas). Naturalmente, esta propriedade pressupõe também que a união numerável $A_1 \cup A_2 \cup A_3 \cup \dots$ está em \mathcal{B} quando cada A_k per-

tença a \mathcal{B} . Nem todas as álgebras de Boole gozam desta propriedade. Aquelas que a possuem dizem-se as σ -álgebras de Boole. Um exemplo é a álgebra de Boole de todos os subconjuntos de S .

DEFINIÇÃO DE PROBABILIDADE PARA CONJUNTOS FUNDAMENTAIS INFINITOS NUMERÁVEIS. *Seja \mathcal{B} uma σ -álgebra de Boole cujos elementos são subconjuntos de um dado conjunto infinito numerável S . Uma função P chama-se uma medida de probabilidade em \mathcal{B} se é não negativa, completamente aditiva, e satisfaz a $P(S) = 1$.*

Quando \mathcal{B} é a álgebra de Boole de todos os subconjuntos de S , fica completamente definida uma função de probabilidade mediante os seus valores em todos os subconjuntos de um só elemento (valores esses que habitualmente se designam por probabilidades pontuais). Todo o subconjunto A de S é finito ou infinito numerável, e a probabilidade de A calcula-se pela adição das probabilidades pontuais para todos os elementos de A ,

$$P(A) = \sum_{x \in A} P(x).$$

A soma no segundo membro é ou uma soma finita, ou uma série infinita absolutamente convergente.

O exemplo que se segue ilustra uma experiência com conjunto fundamental infinito numerável.

EXEMPLO: Lança-se ao ar repetidamente uma moeda até que o primeiro resultado obtido ocorra numa segunda vez; quando tal se verificar termina o jogo.

Para conjunto fundamental tomamos a família de todas as hipóteses de jogos que podem ocorrer. Este conjunto pode exprimir-se como a união de dois conjuntos infinitos numeráveis A e B , onde

$$A = \{TT, THT, THHT, THHHT, \dots\} \quad \text{e} \quad B = \{HH, HTH, HTTH, HTTTH, \dots\}$$

Representamos os elementos do conjunto A (na ordem apresentada) por a_0, a_1, a_2, \dots e os do conjunto B por b_0, b_1, b_2, \dots . Podemos atribuir arbitrariamente probabilidades pontuais não negativas $P(a_n)$ e $P(b_n)$, desde que

$$\sum_{n=0}^{\infty} P(a_n) + \sum_{n=0}^{\infty} P(b_n) = 1.$$

Por exemplo suponhamos que a moeda tem a probabilidade p de sair a face “cara” (H) e probabilidade $q = 1 - p$ de sair a face “cruz” (T), com $0 < p < 1$. Então seria natural a atribuição de probabilidades pontuais

$$(13.26) \quad P(a_n) = q^2 p^n \quad \text{e} \quad P(b_n) = p^2 q^n.$$

Tal atribuição é aceitável porque temos

$$\sum_{n=0}^{\infty} P(a_n) + \sum_{n=0}^{\infty} P(b_n) = q^2 \sum_{n=0}^{\infty} p^n + p^2 \sum_{n=0}^{\infty} q^n = \frac{q^2}{1-p} + \frac{p^2}{1-q} = q + p = 1.$$

Suponhamos agora que queremos saber a probabilidade de que o jogo termine exactamente ao fim de $n+2$ lançamentos. Este é o acontecimento $\{a_n\} \cap \{b_n\}$, e a sua probabilidade é

$$P(a_n) + P(b_n) = q^2 p^n + p^2 q^n.$$

A probabilidade de que termine o jogo quando muito ao fim de $n+2$ lançamentos é

$$\sum_{k=0}^n P(a_k) + \sum_{k=0}^n P(b_k) = q^2 \frac{1-p^{n+1}}{1-p} + p^2 \frac{1-q^{n+1}}{1-q} = 1 - qp^{n+1} - pq^{n+1}.$$

13.22 Exercícios

Os exercícios desta secção referem-se ao exemplo da Secção 13.21.

- Usando as probabilidades pontuais atribuídas na equação (13.26), seja $f_n(p)$ a probabilidade de que o jogo termine exactamente depois de $n+2$ lançamentos. Calcular os valores máximo e mínimo absolutos de $f_n(p)$ no intervalo $0 \leq p \leq 1$ para cada um dos valores $n = 0, 1, 2, 3$.
- Mostrar que é aceitável cada uma das seguintes definições de probabilidades pontuais

$$(a) P(a_n) = P(b_n) = \frac{1}{2^{n+2}} \quad \text{para } n = 0, 1, 2, \dots$$

$$(b) P(a_n) = P(b_n) = \frac{1}{(n+2)(n+3)} \quad \text{para } n = 0, 1, 2, \dots$$

- Calcular a probabilidade de que termine o jogo antes do quinto lançamento, utilizando:
 - as probabilidades pontuais (13.26)
 - as probabilidades pontuais do Exercício 2(a).
 - as probabilidades pontuais do Exercício 2(b).
- Calcular a probabilidade de que seja necessário um número ímpar de lançamentos para terminar o jogo, utilizando:
 - as probabilidades pontuais (13.26).
 - as probabilidades pontuais do Exercício 2(a).
 - as probabilidades pontuais do Exercício 2(b).

13.23 Exercícios variados sobre probabilidades

- Qual a probabilidade de obter uma soma de pontos igual a dez no lançamento de dois dados perfeitos?
- Dez homens e as respectivas esposas estão sentados ao acaso num banquete. Calcular a probabilidade de que um determinado homem esteja sentado junto à sua esposa se: (a) estão sentados a uma mesa redonda; (b) estão sentados em fila.

3. Uma caixa tem duas gavetas. A gaveta número 1 contém quatro moedas de ouro e duas de prata. A gaveta número 2 contém três moedas de ouro e três de prata. Abre-se uma gaveta ao acaso e dela extrai-se ao acaso uma moeda. Calcular a probabilidade de cada um dos seguintes acontecimentos:
 - (a) Que seja aberta a gaveta número 2 e dela se extraia uma moeda de prata.
 - (b) Que seja extraído uma moeda de ouro da gaveta aberta.
4. De um baralho de 52 cartas extraem-se sucessivamente 2 cartas, tendo todas as cartas igual probabilidade de serem escolhidas.
 - (a) Qual a probabilidade de obter pelo menos uma carta de espadas.
Sem examinar as cartas extraídas, lançam-se num saco. Extrai-se do saco uma das cartas e verifica-se que não é espadas.
 - (b) Qual é agora a probabilidade de obter pelo menos uma espada?
A carta previamente extraída do saco é aí colocada de novo e misturam-se as duas cartas. Extrai-se de novo uma carta a qual se examina. Não se faz qualquer comparação para ver se a carta é a mesma que antes de extrair. A carta é de novo colocada no saco e misturam-se. Repete-se esta operação três vezes, incluindo a da alínea (b) e de cada vez a carta examinada verifica-se não ser uma espada.
 - (c) Qual é o conjunto fundamental e a função de probabilidade para esta experiência? Qual é a probabilidade de que uma das duas cartas originais seja uma espada?
5. Um homem tem dez moedas de escudo, nove delas normais e uma com duas faces "cara". Tira uma moeda ao acaso, lança-a ao ar seis vezes e obtém sempre a face cara. Calcular a probabilidade de que a moeda escolhida seja a que tem dupla face cara.
6. Provar que é impossível "viciar" um par de dados de maneira que todo o resultado de obtenção de soma de pontos de dois a doze tenha a mesma probabilidade de ocorrência.
7. Um estudante do segundo ano de Caltech tem um relógio despertador que despertará à hora indicada com uma probabilidade de 0,7. Se desperta, o estudante terá uma probabilidade de 0,8 para chegar a tempo à sua aula de matemática. Se não desperta, a probabilidade de chegar a tempo à sua aula é de 0,3. Calcular a probabilidade de que o estudante chegue a tempo à sua aula de matemática.
8. Três cavalos A , B e C estão numa corrida. O acontecimento " A vence B " será simbolicamente representado por AB . O acontecimento " A vence B , o qual vence C " será representado por ABC , etc. Admitamos que se sabe que

$$P(AB) = \frac{2}{3}, \quad P(AC) = \frac{2}{3}, \quad P(BC) = \frac{1}{2},$$

e que

$$P(ABC) = P(ACB), \quad P(BAC) = P(BCA), \quad P(CAB) = P(CBA).$$

- (a) Calcular a probabilidade de que A vença.
 - (b) Calcular a probabilidade de que B vença.
 - (c) Calcular a probabilidade de que C vença.
 - (d) Serão independentes os acontecimentos AB , AC e CB ?
9. A fase final de um extenso cálculo exige a adição de três inteiros a_1, a_2, a_3 . Suponhamos que:
 - (a) os cálculos de a_1, a_2 e a_3 são aleatoriamente independentes; (b) no cálculo de cada a_i , existe uma mesma probabilidade p de que seja correcto e a probabilidade de cometer um erro igual a $+1$ é a mesma de cometer uma erro igual a -1 ; (c) não se pode cometer um erro maior que $+1$ ou menor que -1 . Tendo presente a possibilidade de compensação dos erros, e calcular a probabilidade de que a soma $a_1 + a_2 + a_3$ esteja correcta.
 10. O jogo de homem desirmanado fora. Suponhamos n pessoas lançando moedas idênticas simultaneamente e independentemente, com $n \geq 3$. Admitamos que é p a probabilidade de obter a face "cara" com cada uma das moedas. Calcular a probabilidade de que, num dado

lançamento, ocorra um "homem desirmanado", isto é, uma pessoa cuja moeda não apresente o mesmo resultado de todas as restantes.

11. Suponhamos que n pessoas jogam ao "homem desirmanado fora" com moedas perfeitas (como se referiu no Exercício 10). Para um dado inteiro m calcular a probabilidade de que sejam necessários precisamente m jogadas para concluir o jogo (isto é que seja na m -ésima jogada que apareça pela primeira vez um homem desirmanado).
12. Suponhamos que uma experiência composta (S, \mathcal{B}, P) é determinada por duas experiências estocasticamente independentes $(S_1, \mathcal{B}_1, P_1)$ e $(S_2, \mathcal{B}_2, P_2)$, onde $S = S_1 \times S_2$ e

$$P(x, y) = P_1(x)P_2(y)$$

para cada (x, y) de S . A finalidade deste exercício é estabelecer a fórmula

$$(13.27) \quad P(U \times V) = P_1(U)P_2(V)$$

para todo o par de subconjuntos U de \mathcal{B}_1 e V de \mathcal{B}_2 . Os espaços amostra S_1 e S_2 supõem-se finitos.

(a) Verificar que a equação (13.27) é verdadeira quando U e V são conjuntos singulares e também quando pelo menos um dos conjuntos U e V é vazio.

Suponhamos agora que

$$U = \{u_1, u_2, \dots, u_k\} \quad \text{e} \quad V = \{v_1, v_2, \dots, v_m\}.$$

Então $U \times V$ consiste de km pares (u_i, v_j) . Para cada $i = 1, 2, \dots, k$ designe A_i o conjunto de m pares de $U \times V$ cujo primeiro elemento é u_i .

- (b) Demonstrar que os A_i são conjuntos disjuntos cuja união é $U \times V$.
- (c) Provar que

$$P(A_i) = \sum_{j=1}^m P(u_i, v_j) = P_1(u_i)P_2(V).$$

- (d) De (b) e (c) deduzir que

$$P(U \times V) = \sum_{i=1}^k P(A_i) = P_1(U)P_2(V).$$

CÁLCULO DAS PROBABILIDADES

14.1 A definição de probabilidade para conjuntos fundamentais não numeráveis

Divide-se um segmento de recta em duas partes, com o ponto de divisão escolhido ao acaso. Qual a probabilidade de que as duas partes tenham o mesmo comprimento? Qual a probabilidade de que a parte maior seja exactamente o dobro da outra? Qual a probabilidade de que a parte maior tenha pelo menos o dobro do comprimento da parte menor? Estes são exemplos de problemas de probabilidades nos quais o conjunto fundamental é não numerável, uma vez que é formado por todos os pontos do segmento de recta. Esta secção estende a definição de probabilidade a conjuntos fundamentais não numeráveis.

Se seguissemos o mesmo processo que para os conjuntos fundamentais numeráveis teríamos que partir com um conjunto não numerável S qualquer e uma σ -álgebra de Boole \mathcal{B} de subconjuntos de S , e definir uma medida de probabilidade que fosse uma função de conjunto P , não negativa, completamente aditiva, definida sobre \mathcal{B} , com $P(S) = 1$. Tal procedimento conduz a certas dificuldades técnicas que não aparecem quando S é numerável. Pretender descrever essas dificuldades levar-nos-ia demasiado longe. Vamos contorná-las impondo restrições iniciais ao conjunto S e à álgebra de Boole \mathcal{B} .

Em primeiro lugar, restringimos S a ser um subconjunto de recta real \mathbf{R} , ou do espaço n -dimensional \mathbf{R}^n . Para a álgebra de Boole \mathcal{B} empregamos subconjuntos especiais de S que, na moderna teoria da integração, são designados por subconjuntos *mensuráveis* de S . Não tentaremos descrever o significado exacto de um conjunto mensurável; em vez disso, mencionaremos algumas das propriedades que possui a classe dos conjuntos mensuráveis.

Consideremos, em primeiro lugar, subconjuntos de \mathbf{R} . Os conjuntos mensuráveis gozam das seguintes propriedades:

1. Se A é mensurável, também o é $\mathbf{R} - A$, o complemento de A .
2. Se $\{A_1, A_2, A_3, \dots\}$ é uma família numerável de conjuntos mensuráveis, então a união $A_1 \cup A_2 \cup A_3 \cup \dots$ é também mensurável.
3. Todo o intervalo (aberto, fechado, semi-aberto, finito ou infinito) é mensurável.

Assim, os conjuntos mensuráveis de \mathbf{R} formam uma σ -álgebra de Boole que contém os intervalos. Existe uma σ -álgebra de Boole mínima, a qual goza desta propriedade; os seus elementos dizem-se *conjuntos de Borel*, em atenção ao matemático francês Émile Borel (1871-1956).

Analogamente, no espaço bidimensional existe uma σ -álgebra de Boole mínima a qual contém todos os produtos cartesianos de pares de intervalos; os seus elementos são os conjuntos de Borel a duas dimensões. Os conjuntos de Borel, no espaço n -dimensional, definem-se de uma maneira análoga.

De aqui em diante, sempre que usemos um conjunto S de números reais como conjunto fundamental, ou, mais geralmente, sempre que usemos um conjunto S do espaço n -dimensional como conjunto fundamental, suporemos que este é um conjunto de Borel. Os subconjuntos de Borel de S formam eles próprios uma σ -álgebra de Boole. Estes subconjuntos são suficientemente genéricos para incluírem todos os acontecimentos que ocorrem nas aplicações correntes da teoria das probabilidades.

DEFINIÇÃO DE PROBABILIDADE PARA CONJUNTOS FUNDAMENTAIS NÃO NUMERÁVEIS. *Seja S um subconjunto de \mathbf{R}^n e \mathcal{B} a σ -álgebra de Boole de subconjuntos de Borel de S . Uma função de conjunto P não negativa, completamente aditiva, definida em \mathcal{B} com $P(S) = 1$ chama-se uma medida de probabilidade. O terno (S, \mathcal{B}, P) chama-se um espaço de probabilidade.*

14.2 Numerabilidade de conjuntos de pontos com probabilidade positiva

Para conjuntos fundamentais numeráveis a probabilidade de um acontecimento A calcula-se muitas vezes somando as probabilidades pontuais $P(x)$ para todo x de A . Este método não é vantajoso para conjuntos fundamentais não numeráveis porque, como mostra o teorema seguinte, a maior parte das probabilidades pontuais são nulas

TEOREMA 14.1. *Se (S, \mathcal{B}, P) é um espaço de probabilidade e se T representa o conjunto de todos os x de S para os quais $P(x) > 0$, então T é numerável.*

Demonstração. Para cada $n = 1, 2, 3, \dots$, seja T_n o seguinte subconjunto de S :

$$T_n = \left\{ x \mid \frac{1}{n+1} < P(x) \leq \frac{1}{n} \right\}.$$

Se $P(x) > 0$, então $x \in T_n$ para algum n . Reciprocamente, se $x \in T_n$ para algum n então $x \in T$. Logo $T = T_1 \cup T_2 \cup \dots$. Agora T_n contém quando muito n pontos, porque se contivesse $n+1$ ou mais pontos a soma das respectivas probabilidades pontuais excederia 1. Portanto T é numerável, visto ser uma união numerável de conjuntos finitos.

O teorema 14.1 diz-nos que podem atribuir-se probabilidades positivas quando muito a um subconjunto numerável de S . Os restantes pontos de S terão probabilidade zero. Em particular, se todos os resultados de S são igualmente prováveis, então a todo o ponto de S deve prescrever-se a probabilidade zero.

Nota: Ao teorema 14.1 pode dar-se uma interpretação física em termos de distribuição de massa probabilística, o que nos permitirá ilustrar o seu significado. Imaginemos que dispomos de uma massa total igual a 1. (Isto corresponde a $P(S) = 1$). Admitamos que podemos distribuir esta massa da maneira que se pretenda ao longo do eixo real, quer de modo uniforme ou variável, ou então colocando porções discretas de massa em certos pontos, ou de ambos os modos. (Interpretamos uma quantidade positiva de massa como uma porção discreta). Podemos colocar toda a massa num ponto. Podemos dividir a massa em porções discretas iguais ou desiguais entre dois pontos, dez pontos, um milhão de pontos, ou entre um conjunto infinito numerável de pontos. Por exemplo, podemos por $\frac{1}{2}$ em 1, $\frac{1}{4}$ em 2, $\frac{1}{8}$ em 3, etc., ou seja pôr a massa $(\frac{1}{2})^n$ em cada inteiro $n \geq 1$. Ou podemos distribuir toda a massa de uma maneira contínua. Ou ainda podemos distribuir continuamente parte dela e dividir o resto em porções discretas. O teorema 14.1 diz-nos que, quando muito, podemos atribuir quantidades discretas de massa a um conjunto numerável de pontos.

Posto que a maior parte (se não todas) das probabilidades pontuais para um conjunto fundamental não numerável serão zero, o conhecimento das probabilidades pontuais por si só não basta para o cálculo das probabilidades de acontecimentos aleatórios. É necessária mais informação; a descrição é mais completa utilizando dois novos conceitos, as *variáveis aleatórias* e as *funções de repartição*, (ou *distribuição*), a que nos vamos referir a seguir. Os referidos conceitos tornam possível o uso do cálculo integral em muitos problemas cujo conjunto fundamental é não numerável. A integração substitui, então, a adição na determinação das probabilidades.

14.3 Variáveis aleatórias

Em muitas experiências estamos interessados em *números* associados com os resultados da experiência. Por exemplo, lança-se ao ar n moedas simultaneamente e pretendemos saber qual o número de vezes que se obtese faces “cara”. Lança-se um par de dados e pede-se a soma dos pontos obtidos. Aponta-se uma flecha num alvo circular e pretendemos saber a sua distância ao centro do alvo. Sempre que associamos um número real a cada resultado de uma experiência estamos a considerar uma *função* cujo domínio é o conjunto dos resultados possíveis e o contradomínio é o conjunto dos números reais em questão. Uma tal função diz-se uma *variável aleatória*. Apresentamos a seguir uma definição formal:

DEFINIÇÃO DE VARIÁVEL ALEATÓRIA. *Seja S o conjunto fundamental. Uma função real definida em S chama-se uma variável aleatória unidimensional. Se os valores da função são pares ordenados de números reais (isto é, vectores no espaço bidimensional), a função diz-se ser uma variável aleatória bidimensional. Mais geralmente, uma variável aleatória n*

dimensional é simplesmente uma função cujo domínio é o conjunto fundamental S e cujo contradomínio é uma família de n -tuplos de números reais (vectors num n espaço).

Assim, uma variável aleatória não é mais do que uma função vectorial definida num conjunto. O termo “aleatória” é usado unicamente para recordar que o conjunto em questão é um conjunto fundamental.†

Devido à generalidade da definição anterior, é possível ter variáveis aleatórias distintas associadas a uma mesma experiência. Em cada caso particular o experimentador deve decidir quais as variáveis aleatórias que lhe interessam. Em geral, procuramos trabalhar com variáveis aleatórias cujos valores reflectam, o mais simplesmente possível, as propriedades dos resultados da experiência que são realmente essenciais.

Notações: As letras maiúsculas tais como X, Y, Z são habitualmente utilizadas para representar variáveis aleatórias unidimensionais. Um resultado típico da experiência (isto é, um elemento genérico do conjunto fundamental) representa-se habitualmente pela letra grega ω . Assim, $X(\omega)$ representa aquele número real que a variável aleatória X associa com o resultado ω .

Apresentamos a seguir alguns exemplos de variáveis aleatórias.

EXEMPLO 1. Uma experiência consiste no lançamento de um dado e na leitura do número de pontos obtidos. A variável aleatória X “mais natural” a considerar é o número de pontos gravados pelo fabricante no dado, a saber:

$$X(\omega) = \omega \quad \text{para} \quad \omega = 1, 2, 3, 4, 5, 6.$$

Se estamos interessados no caso em que o número de pontos é par ou ímpar, então podemos agora considerar outra variável aleatória Y , definida do modo seguinte:

$$Y(\omega) = 0 \quad \text{se } \omega \text{ é par,}$$

$$Y(\omega) = 1 \quad \text{se } \omega \text{ é ímpar.}$$

Os valores 0 e 1 não são essenciais—poderia em seu lugar usar-se qualquer par de números reais distintos. Contudo 0 e 1 sugerem “par” ou “ímpar”, respectivamente, porque representam o resto que se obtém quando o resultado ω se divide por 2.

EXEMPLO 2. Aponta-se uma flecha num alvo circular. O conjunto de todos os resultados possíveis é o conjunto de todos os pontos ω do alvo. Se imaginamos um sistema de eixos coordenados com a origem no centro do alvo, podemos definir várias variáveis aleatórias para esta experiência. Uma delas é a variável aleatória bidi-

† A expressão “variável estocástica” também se usa como sinónimo de “variável aleatória”. A palavra estocástica procede de uma raiz grega e significa “sorte” e segundo parece foi introduzida por Bernoulli e é de uso frequente na teoria das probabilidades.

mensional que atribui ao ponto ω as suas coordenadas rectangulares (x, y) . Outra é a que atribui a ω as coordenadas polares r e θ . Exemplos de variáveis aleatórias unidimensionais são aquelas que atribuem a cada ω precisamente uma das suas coordenadas, tais como por exemplo x ou r . Numa experiência deste tipo pretendemos muitas vezes saber a probabilidade de que a flecha atinja determinada região do alvo, por exemplo o primeiro quadrante. Este acontecimento pode descrever-se muito simplesmente por intermédio da variável aleatória que atribui a cada ponto ω a sua coordenada polar θ , de modo que $X(\omega) = \theta$; o acontecimento “a flecha no primeiro quadrante” é o conjunto de pontos ω tais que $0 \leq X(\omega) \leq \frac{1}{2}\pi$.

Abreviaturas. Evitamos notações embaraçosas recorrendo a abreviaturas especiais para descrever certos tipos de acontecimentos e as respectivas probabilidades. Por exemplo, se t é um número real, o conjunto de todos os ω do conjunto fundamental tais que $X(\omega) = t$ representa-se abreviadamente por

$$X = t.$$

A probabilidade deste acontecimento escreve-se $P(X = t)$ em vez de $P(\{\omega \mid X(\omega) = t\})$. Símbolos tais como $P(X = a \text{ ou } X = b)$ e $P(a < X \leq b)$ definem-se de uma maneira semelhante. Assim, o acontecimento “ $X = a$ ou $X = b$ ” é a união dos dois acontecimentos “ $X = a$ e $X = b$ ”; o símbolo $P(X = a \text{ ou } X = b)$ representa a probabilidade desta união. O acontecimento “ $a < X \leq b$ ” é o conjunto de todos os pontos ω tais que $X(\omega)$ pertence ao intervalo semi aberto $(a, b]$, e o símbolo $P(a < X \leq b)$ representa a probabilidade desse acontecimento.

14.4. Exercícios

- Seja X uma variável aleatória unidimensional.
 - Se $a < b$, mostrar que os dois acontecimentos $a < X \leq b$ e $X \leq a$ são disjuntos.
 - Determinar a união dos dois acontecimentos na alínea (a).
 - Provar que $P(a < X \leq b) = P(X \leq b) - P(X \leq a)$.
- Seja (X, Y) uma variável aleatória bidimensional definida sobre o conjunto fundamental S . Quer isto dizer que (X, Y) é uma função que atribui a cada ω de S um par de números reais $(X(\omega), Y(\omega))$. Evidentemente, X e Y são variáveis aleatórias unidimensionais definidas em S . A notação

$$X \leq a, Y \leq b$$

representa o conjunto de todos os elementos ω de S tais que $X(\omega) \leq a$ e $Y(\omega) \leq b$.

- Se $a < b$ e $c < d$ descrever, em função dos elementos de S , o significado da seguinte notação: $a < X \leq b, c < Y \leq d$.
- Mostrar que os dois acontecimentos “ $X \leq a, Y \leq c$ ” e “ $X \leq a, c < Y \leq d$ ” são disjuntos. Interpretar estes acontecimentos geometricamente.
- Determinar a união dos dois acontecimentos em (b).
- Generalizar o Exercício 1 (c) ao caso bidimensional.

3. Lançam-se dois dados perfeitos, sendo o resultado de cada lançamento expresso por um par ordenado (a, b) , em que cada um dos valores a e b é um inteiro de 1 a 6. Seja X a variável aleatória que atribui o valor $a + b$ ao resultado (a, b) .
 - (a) Descrever as hipóteses de realização dos acontecimentos " $X = 7$ ", " $X = 11$ ", " $X = 7$ ou $X = 11$ ".
 - (b) Calcular as probabilidades dos acontecimentos da alínea (a).
4. Considere-se uma experiência que consiste no lançamento simultâneo de quatro moedas (ou no lançamento de uma moeda quatro vezes). Para cada moeda definir uma variável aleatória que atribui o valor 1 à saída da face "cara" e o valor 0 à saída da face "cruz", e representem-se essas variáveis aleatórias por X_1, X_2, X_3, X_4 , atribuindo as probabilidades $P(X_i = 1) = P(X_i = 0) = \frac{1}{2}$ para cada X_i . Considerar uma nova variável aleatória Y que atribui a cada resultado o número total de faces "cara" obtido nas quatro moedas. Expressar Y em função de X_1, X_2, X_3, X_4 e calcular as probabilidades $P(Y = 0)$, $P(Y = 1)$, e $P(Y \leq 1)$.
5. Uma pequena empresa ferroviária tem meios apenas para o transporte de 100 passageiros por dia entre duas cidades, a um preço fixo (para a companhia) de 7\$00 por passageiro. Se um dia compram bilhetes mais de 100 passageiros a companhia é obrigada a proporcionar o transporte em autocarro aos passageiros que excedem a centena ao preço de 10\$00 por pessoa. Seja X a variável aleatória que representa o número de passageiros que compram bilhetes num determinado dia. Os valores possíveis de X são os inteiros 0, 1, 2, 3, ..., até um certo máximo desconhecido. Se Y representa a variável aleatória que exprime o custo total diário (em escudos) para a companhia pelos passageiros transportados, exprimir Y em função de X .
6. Uma fábrica dotada de linha de produção em cadeia tem dois terminais A e B . Em A montam-se X unidades por hora; estas são imediatamente transportadas para B , onde são inspeccionadas à razão de Y unidades por hora, sendo $Y < X$. Os valores possíveis de X e Y são os inteiros 8, 9 e 10. Seja Z a variável aleatória que representa o número de unidades que saem da linha de produção durante a primeira hora.
 - (a) Expressar Z em função de X e Y supondo que cada uma das variáveis X e Y é constante durante a primeira hora.
 - (b) Descrever, de maneira análoga, a variável aleatória U que dá conta do número de unidades produzidas nas duas primeiras horas consecutivas de produção. Cada uma das variáveis X e Y é constante durante cada hora, mas os valores constantes durante a segunda hora não coincidem necessariamente com os da primeira hora.

14.5. Funções de repartição

Voltamos agora de novo ao problema do cálculo de probabilidades de acontecimentos associados com uma dada variável aleatória. Seja X uma variável aleatória unidimensional definida num conjunto fundamental S , sendo S um conjunto de Borel num n -espaço para $n \geq 1$. Seja P uma medida de probabilidade definida nos subconjuntos de Borel de S . Para cada ω de S , $X(\omega)$ é um número real, e quando ω percorre os elementos de S os números $X(\omega)$ percorrem um conjunto de números reais (0 contra domínio de X). Este conjunto pode ser finito, infinito numerável, ou não numerável. Para cada real t consideramos o seguinte subconjunto particular de S :

$$A(t) = \{\omega \mid X(\omega) \leq t\}.$$

Se t é menor que todos os números do contradomínio de X , o conjunto $A(t)$ será vazio; caso contrário, $A(t)$ será um subconjunto não vazio de S . Admitamos que para cada t o conjunto $A(t)$ é um acontecimento, isto é, um conjunto de Borel. De acordo com o estabelecido na final da Secção 14.3, representamos este acontecimento pelo símbolo $X \leq t$.

Suponhamos conhecida a probabilidade $P(X \leq t)$ para todo o real t . Veremos que o conhecimento desta probabilidade permitir-nos-á calcular as probabilidades de muitos outros acontecimentos de interesse. Consegue-se isso usando as probabilidades $P(X \leq t)$ como base para a construção de uma nova função F , chamada a *função de distribuição* ou de *repartição* de X , que se define da maneira seguinte:

DEFINIÇÃO DA FUNÇÃO DE REPARTIÇÃO. *Seja X uma variável aleatória unidimensional. A função F definida para todo o real t pela equação*

$$F(t) = P(X \leq t)$$

chama-se a função de repartição da variável aleatória X .

Nota: Algumas vezes, utiliza-se a notação F_X para evidenciar o facto de que a função de repartição está associada a uma certa variável aleatória X . O valor da função em t escreve-se então $F_X(t)$.

É importante ter presente que a função de repartição F está definida ao longo de todo o eixo real, muito embora o contradomínio de X possa ser unicamente uma parte limitada desse eixo real. Com efeito, se todos os números $X(\omega)$ estão num certo intervalo finito $[a, b]$, então para $t < a$ a probabilidade $P(X \leq t)$ é zero (visto que para $t < a$ o conjunto $X \leq t$ é vazio) e para $t \geq b$ a probabilidade $P(X \leq t)$ é 1 (porque neste caso o conjunto $X \leq t$ é todo o conjunto fundamental). Isto significa que para variáveis aleatórias X limitadas cujo contradomínio é interior ao intervalo $(a, b]$ se tem $F(t) = 0$ para todo $t < a$ e $F(t) = 1$ para todo $t \geq b$.

Passamos em seguida ao estabelecimento de um certo número de propriedades comuns a todas as funções de repartição.

TEOREMA 14.2. *Se F é uma função de repartição de uma variável aleatória unidimensional X , então tem-se:*

- (a) $0 \leq F(t) \leq 1$ para todo t .
- (b) $P(a < X \leq b) = F(b) - F(a)$ se $a < b$.
- (c) $F(a) \leq F(b)$ se $a < b$.

Demonstração: A alínea (a) resulta imediatamente da definição de F porque as probabilidades estão sempre compreendidas entre 0 e 1.

Para demonstrar (b) observe-se que os acontecimentos " $a < X \leq b$ " e " $X \leq a$ " são disjuntos. A sua união é o acontecimento " $X \leq b$ ". Usando a aditividade obtemos

$$P(a < X \leq b) + P(X \leq a) = P(X \leq b),$$

que também pode pôr-se na forma

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a).$$

A alínea (c) resulta de (b), uma vez que $P(a < X \leq b) \geq 0$.

Nota: Usando a analogia da distribuição de massa, podemos dizer que $F(t)$ representa a quantidade de massa total situada entre $-\infty$ e t (incluindo o próprio ponto t). O total de massa localizada no intervalo semi-aberto $(a, b]$ é $F(b) - F(a)$.

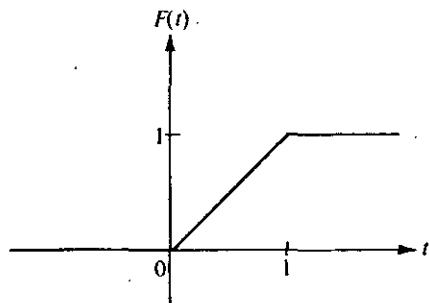


FIG. 14.1. Função de repartição de uma variável aleatória limitada.

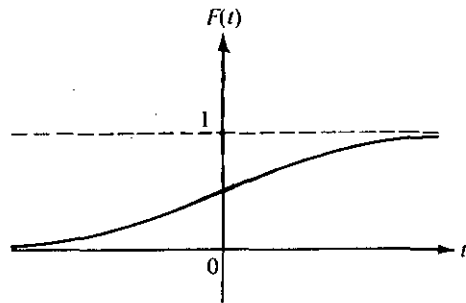


FIG. 14.2. Função de repartição de uma variável aleatória não limitada.

A figura 14.1 mostra uma função de repartição de uma variável aleatória X cujos valores $X(\omega)$ estão no intervalo $(0, 1]$. Este exemplo particular é conhecido como uma *distribuição uniforme*. Aqui tem-se

$$F(t) = 0 \text{ para } t < 0, \quad F(t) = t \text{ para } 0 \leq t \leq 1, \quad F(t) = 1 \text{ para } t \geq 1.$$

A figura 14.2 mostra um exemplo de uma função de repartição correspondendo a uma variável aleatória não limitada. Este exemplo é conhecido por *distribuição de Cauchy* e os valores correspondentes são dados por

$$F(t) = \frac{1}{2} + \frac{1}{\pi} \arctan t.$$

Mais adiante analisaremos casos práticos que conduzem a distribuições uniformes e de Cauchy.

Nota: Recorrendo à analogia da distribuição de massa, poderíamos dizer que na figura 14.1 não foi distribuída qualquer quantidade de massa probabilística à esquerda da origem ou à direita do ponto 1. Toda a massa foi distribuída no intervalo $[0, 1]$. O gráfico de F , neste intervalo, é uma recta porque a massa foi distribuída com uma densidade constante. Na figura 14.2 a massa probabilística foi

distribuída ao longo de todo o eixo. O gráfico não é uma recta porque a distribuição se fez com densidade variável.

O teorema 14.2 (b) diz-nos como calcular (em função de F) a probabilidade de que X esteja num intervalo semi-aberto da forma $(a, b]$. O teorema que apresentamos a seguir refere-se a outros tipos de intervalos.

TEOREMA 14.3. *Seja F uma função de repartição de uma variável aleatória unidimensional X . Então se $a < b$ tem-se*

- (a) $P(a \leq X \leq b) = F(b) - F(a) + P(X = a)$.
- (b) $P(a < X < b) = F(b) - F(a) - P(X = b)$.
- (c) $P(a \leq X < b) = F(b) - F(a) + P(X = a) - P(X = b)$.

Demonstração. Para demonstrar (a) observemos que os acontecimentos " $a < X \leq b$ " e " $X = a$ " são disjuntos e a sua união é " $a \leq X \leq b$ ". Utilizando a aditividade e o teorema 14.2 (b) obtemos (a). As alíneas (b) e (c) demonstram-se do mesmo modo.

Deve notar-se que os quatro acontecimentos

$$a < X \leq b, \quad a \leq X \leq b, \quad a < X < b, \quad \text{e} \quad a \leq X < b$$

têm igual probabilidade se e só se $P(X = a) = 0$ e $P(X = b) = 0$.

Os exemplos representados na figura 14.1 e 14.2 evidenciam-se com duas outras propriedades inerentes a todas as funções de repartição, as quais se enunciam no teorema seguinte.

TEOREMA 14.4. *Se F é a função de repartição de uma variável aleatória unidimensional X então tem-se*

$$(14.1) \quad \lim_{t \rightarrow -\infty} F(t) = 0 \quad \text{e} \quad \lim_{t \rightarrow +\infty} F(t) = 1.$$

Demonstração. A existência dos dois limites em (14.1) e o facto de que os dois limites estão entre 0 e 1 é de conclusão imediata, visto que F é uma função monótona cujos valores estão compreendidos entre 0 e 1.

Representamos os limites em (14.1) respectivamente por L_1 e L_2 . Para demonstrarmos que $L_1 = 0$ e que $L_2 = 1$ recorreremos à propriedade da probabilidade ser completamente aditiva. Para isso exprimimos todo o espaço S como uma união numerável de acontecimentos disjuntos:

$$S = \bigcup_{n=1}^{\infty} (-n < X \leq -n+1) \cup \bigcup_{n=0}^{\infty} (n < X \leq n+1).$$

Então, pela aditividade, obtemos

$$\begin{aligned}
 P(S) &= \sum_{n=1}^{\infty} P(-n < X \leq -n+1) + \sum_{n=0}^{\infty} P(n < X \leq n+1) \\
 &= \lim_{M \rightarrow \infty} \sum_{n=1}^M [F(-n+1) - F(-n)] + \lim_{N \rightarrow \infty} \sum_{n=0}^N [F(n+1) - F(n)].
 \end{aligned}$$

Várias parcelas do segundo membro anulam-se mutuamente, dando-nos

$$\begin{aligned}
 P(S) &= \lim_{M \rightarrow \infty} [F(0) - F(-M)] + \lim_{N \rightarrow \infty} [F(N+1) - F(0)] \\
 &= F(0) - L_1 + L_2 - F(0) = L_2 - L_1.
 \end{aligned}$$

Porque $P(S) = 1$, a igualdade anterior significa que $L_2 - L_1 = 1$ ou seja que $L_2 = 1 + L_1$. Por outro lado, temos também que $L_2 \leq 1$ e $L_1 \geq 0$. Isto implica que $L_2 = 0$ e $L_1 = 1$, como se afirmara.

14.6. Discontinuidades das funções de repartição

Um exemplo de possibilidade de existência de uma função de repartição com discontinuidades é apresentada na figura 14.3. Recorrendo mais uma vez à analogia de distribuição de massa poderíamos dizer que F tem uma discontinuidade (a que cor-

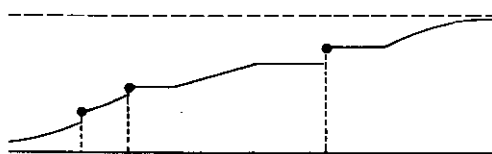


FIG. 14.3. Uma função de repartição possível.

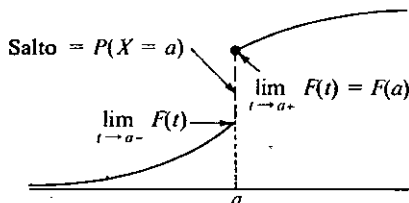


FIG. 14.4. Discontinuidade (em salto ou degrau) de uma função de repartição.

responde um salto ou degrau) em cada ponto que contém uma quantidade positiva de massa. Como o teorema seguinte mostra, o salto é igual ao total de massa probabilística concentrada naquele ponto particular.

TEOREMA 14.5. *Se F é a função de repartição de uma variável aleatória unidimensional X , então para cada valor real t tem-se*

$$(14.2) \quad \lim_{t \rightarrow a+} F(t) = F(a)$$

e

$$(14.3) \quad \lim_{t \rightarrow a-} F(t) = F(a) - P(X = a).$$

Nota: O limite (14.2) diz-nos que F é contínua à direita de cada ponto a , porque $F(t) \rightarrow F(a)$ quando $t \rightarrow a$ por valores à direita de a . Por outro lado (14.3) diz-nos que quando $t \rightarrow a$ por valores à esquerda de a , $F(t)$ tenderá para $F(a)$ se e só se a probabilidade $P(X = a)$ é zero. Quando $P(X = a)$ não é zero, o gráfico de F apresenta um salto em a do tipo representado na figura 14.4.

Demonstração. A existência dos limites resulta imediatamente de F ser monótona e limitada. Interessa agora provar que os limites têm os valores indicados. Para tal consideremos a alínea (b) do teorema 14.2. Se $t > a$ escrevemos

$$(14.4) \quad F(t) = F(a) + P(a < X \leq t);$$

se $t < a$ escrevemos

$$(14.5) \quad F(t) = F(a) - P(t < X \leq a).$$

Fazendo $t \rightarrow a+$ em (14.4) encontramos

$$\lim_{t \rightarrow a+} F(t) = F(a) + \lim_{t \rightarrow a+} P(a < X \leq t),$$

enquanto que se $t \rightarrow a-$ em (14.5) obtemos

$$\lim_{t \rightarrow a-} F(t) = F(a) - \lim_{t \rightarrow a-} P(t < X \leq a).$$

Portanto para demonstrar (14.2) e (14.3) devemos estabelecer duas igualdades:

$$(14.6) \quad \lim_{t \rightarrow a+} P(a < X \leq t) = 0$$

e

$$(14.7) \quad \lim_{t \rightarrow a-} P(t < X \leq a) = P(X = a).$$

Estas podem justificar-se intuitivamente do modo seguinte: quando $t \rightarrow a+$, o intervalo semi-aberto $(a, t]$ tende a confundir-se com o intervalo vazio, isto é, a intersecção de todos os intervalos semi-abertos $(a, t]$ para $t > a$, é vazia. Por outro lado, quando $t \rightarrow a-$ o intervalo semi-aberto $(t, a]$ tende a confundir-se com o ponto a . (A intersecção de todos os intervalos $(t, a]$ para $t > a$ é o conjunto $\{a\}$.) Deste modo, se a probabilidade se comporta de maneira contínuas as igualdades (14.6) e (14.7) devem ser válidas. Para conduzir este raciocínio a uma demonstração rigorosa procedemos do modo seguinte:

Para todo inteiro $n \geq 1$, seja

$$(14.8) \quad p_n = P\left(a < X \leq a + \frac{1}{n}\right).$$

Para provar (14.6) basta demonstrar aqui que $p_n \rightarrow 0$ quando $n \rightarrow \infty$. Representemos por S_n o acontecimento

$$a + \frac{1}{n+1} < X \leq a + \frac{1}{n}.$$

Os conjuntos S_n são disjuntos e a sua reunião $S_1 \cup S_2 \cup S_3 \cup \dots$ é o acontecimento $a < X \leq a + 1$. Pela aditividade numerável temos:

$$(14.9) \quad \sum_{n=1}^{\infty} P(S_n) = P(a < X \leq a + 1) = p_1.$$

Por outro lado, a equação (14.8) implica que

$$p_n - p_{n+1} = P(S_n),$$

pelo que de (14.9) obtemos a relação

$$(14.10) \quad \sum_{n=1}^{\infty} (p_n - p_{n+1}) = p_1.$$

A convergência da série é uma consequência de (14.9). Mas a série no primeiro membro de (14.10) tem por soma

$$p_1 - \lim_{n \rightarrow \infty} p_n.$$

Por conseguinte (14.10) implica que $\lim_{n \rightarrow \infty} p_n = 0$, o que prova (14.6).

Uma ligeira modificação deste raciocínio permite-nos demonstrar igualmente (14.7). Visto que

$$P(t < X \leq a) = P(t < X < a) + P(X = a)$$

necessitamos provar unicamente que

$$\lim_{t \rightarrow a-} P(t < X < a) = 0.$$

Com tal finalidade introduzimos os números

$$q_n = P\left(a - \frac{1}{n} < X < a\right)$$

e provamos que $q_n \rightarrow 0$ quando $n \rightarrow \infty$. Neste caso consideremos os acontecimentos T_n definidos por

$$a - \frac{1}{n} < X \leq a - \frac{1}{n+1}$$

para $n = 1, 2, 3, \dots$. Estes são acontecimentos disjuntos e a respectiva união é o acontecimento $a - 1 < X < a$, pelo que se tem

$$\sum_{n=1}^{\infty} P(T_n) = P(a - 1 < X < a) = q_1.$$

Tendo presente que $q_n - q_{n+1} = P(T_n)$, completamos a demonstração como atrás.

O tipo mais geral de função de repartição é qualquer função real F gozando das propriedades seguintes:

- (a) F é monótona crescente em todo o eixo real,
- (b) F é contínua à direita de cada ponto,
- (c) $\lim_{t \rightarrow -\infty} F(t) = 0$ e $\lim_{t \rightarrow +\infty} F(t) = 1$.

Com efeito pode demonstrar-se que, para cada tal função F , existe uma correspondente função de conjunto P , definida nos conjuntos de Borel do eixo real, tal que F é uma medida de probabilidade que atribui a probabilidade $F(b) - F(a)$ a cada intervalo semi-aberto $(a, b]$. Para uma demonstração desta proposição ver H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, N. J.: 1946.

Existem dois tipos especiais de distribuições, chamados *discretas* e *contínuas*, que assumem particular importância na prática. No caso discreto toda a massa está concentrada num número finito, ou numa infinidade numerável de pontos, enquanto que no caso contínuo a massa é distribuída, com densidade constante ou variável, ao longo de um intervalo (finito ou infinito). Estes dois tipos de distribuições serão tratados em pormenor nas próximas secções.

14.7. Distribuições discretas. Funções de massa probabilística

Seja X uma variável aleatória unidimensional e consideremos uma nova função p chamada a *função massa probabilística* de X . Os seus valores $p(t)$ são definidos, para cada número real t , pela equação.

$$p(t) = P(X = t).$$

Quer dizer, $p(t)$ é a probabilidade de que X tome o valor de t . Quando pretendemos pôr em destaque que p está associada com X escrevemos p_X em vez de p e $p_X(t)$ em vez de $p(t)$.

O conjunto dos números reais t para os quais $p(t) > 0$ é finito ou numerável; representamo-lo por T e será

$$T = \{t \mid p(t) > 0\}.$$

A variável aleatória X diz-se *discreta* se

$$\sum_{t \in T} p(t) = 1.$$

Por outras palavras, X é discreta se a unidade de massa probabilística está distribuída sobre o eixo real, concentrando-se uma massa positiva $p(t)$ em cada ponto t de um certo conjunto T , finito ou infinito numerável, e não existindo qualquer massa nos restantes pontos. Os pontos de T chamam-se os *pontos de massa probabilística* de X .

Para variáveis aleatórias discretas o conhecimento da função de massa probabilística permite-nos calcular as probabilidades de acontecimentos arbitrários. Com efeito, existe o seguinte:

TEOREMA 14.6. *Se A é um subconjunto de Borel da recta real \mathbf{R} , e se $P(X \in A)$ representa a probabilidade do conjunto de ω tal que $X(\omega) \in A$, então*

$$(14.11) \quad P(X \in A) = \sum_{x \in A \cap T} p(x),$$

onde T é o conjunto de pontos de massa probabilística de X .

Demonstração. Uma vez que $A \cap T \subseteq A$ e $T - A \subseteq \mathbf{R} - A$, temos

$$(14.12) \quad \sum_{x \in A \cap T} p(x) \leq P(X \in A) \quad \text{e} \quad \sum_{x \in T - A} p(x) \leq P(X \notin A).$$

Mas $A \cap T$ e $T - A$ são conjuntos disjuntos cuja união é T , pelo que a segunda desigualdade, em (14.12) é equivalente a

$$1 - \sum_{x \in A \cap T} p(x) \leq 1 - P(X \in A) \quad \text{ou} \quad \sum_{x \in A \cap T} p(x) \geq P(X \in A).$$

Combinando esta com a primeira desigualdade de (14.12) obtemos (14.11).

Nota: Uma vez que $p(x) = 0$ quando $x \notin T$, a soma do segundo membro de (14.11) pode escrever-se $\sum_{x \in A} p(x)$ sem perigo de confusão.

Quando A é o intervalo $(-\infty, t]$, a soma do segundo membro de (14.11) dá o valor da função de repartição $\bar{F}(t)$. Assim, temos

$$\bar{F}(t) = P(X \leq t) = \sum_{x \leq t} p(x).$$

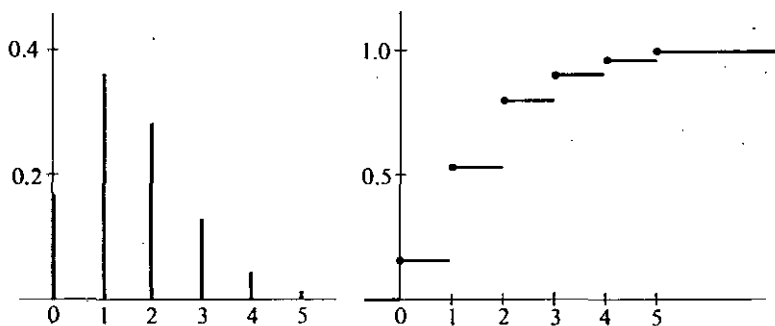
Se uma variável aleatória X é discreta, a correspondente função de repartição F diz-se também discreta.

Na prática ocorrem frequentemente os seguintes exemplos de distribuições discretas.

EXEMPLO 1. Distribuição binomial. Seja p um número real dado satisfazendo a $0 \leq p \leq 1$ e $q = 1 - p$. Suponhamos uma variável aleatória X que toma os valores 0,

$1, 2, \dots, n$ com n um inteiro positivo fixo, e suponhamos que a probabilidade $P(X = k)$ é dada por

$$P(X = k) = \binom{n}{k} p^k q^{n-k} \quad \text{para } k = 0, 1, 2, \dots, n.$$



(a) A função massa probabilística

(b) A função de repartição.

FIG. 14.5. A função massa probabilística e a função de repartição de uma distribuição binomial com parâmetros $n = 5$ e $p = \frac{1}{3}$.

Esta definição de probabilidades pontuais está correcta porque a soma de todas elas é

$$\sum_{k=0}^n P(X = k) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = (p + q)^n = 1.$$

A correspondente função de repartição F_X diz-se ser uma *distribuição binomial* de parâmetros n e p . Os seus valores podem calcular-se pela fórmula

$$F_X(t) = \sum_{0 \leq k \leq t} \binom{n}{k} p^k q^{n-k}.$$

As distribuições binomiais apresentam-se, de modo natural, no caso de uma sucessão de experiências num esquema de Bernoulli, sendo p a probabilidade do “acontecimento favorável” e $q = 1 - p$ a probabilidade do “acontecimento contrário”. Na verdade, quando a variável aleatória X dá conta do número de acontecimentos favoráveis em n experiências, $P(X = k)$ é precisamente $\binom{n}{k} p^k q^{n-k}$ em virtude da fórmula de Bernoulli. (Ver o teorema 13.3 da Secção 13.16). A figura 14.5 mostra o gráfico da função massa probabilística e a correspondente função de repartição para uma distribuição binomial com os parâmetros $n = 5$ e $p = \frac{1}{3}$.

EXEMPLO 2. Distribuição de Poisson. Seja λ um número real e positivo e X uma variável aleatória que toma os valores $0, 1, 2, 3, \dots$. Se a probabilidade $P(X = k)$ é dada por

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!} \quad \text{para } k = 0, 1, 2, \dots,$$

a correspondente função de repartição F_X diz-se definir uma *distribuição de Poisson* com o parâmetro λ . É assim designada em homenagem ao matemático francês S. D. Poisson (1781-1840). É válida esta definição de probabilidades porque

$$\sum_{k=0}^{\infty} P(X = k) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1.$$

Os valores da função de repartição calculam-se através das somas parciais

$$F_X(t) = e^{-\lambda} \sum_{0 \leq k \leq t} \frac{\lambda^k}{k!}.$$

A distribuição de Poisson é aplicável a muitos problemas relativos a acontecimentos aleatórios ocorrendo no tempo, tais como acidentes de tráfego, ligações para números errados em comunicações telefónicas, e trocas de cromossomas em células provocadas por raios X . Nos livros de Feller e Parzen, citados no final deste capítulo, faz-se a análise de algumas aplicações específicas.

14.8. Exercícios

1. Lança-se um dado perfeito. Para variável aleatória X toma-se o número de pontos obtidos. Desenhar o gráfico da correspondente função de repartição F_X .
2. Lançam-se simultaneamente dois dados perfeitos. Seja X a variável aleatória definida pelo total de pontos obtidos. Construir uma tabela para os valores não nulos de função massa probabilística p_X e traçar o gráfico da correspondente função de repartição F_X .
3. A função de repartição F de uma variável aleatória X define-se do modo seguinte:

$$F(t) = \begin{cases} 0 & \text{se } t < -2, \\ \frac{1}{2} & \text{se } -2 \leq t < 0, \\ \frac{3}{4} & \text{se } 0 \leq t < 2, \\ 1 & \text{se } t \geq 2. \end{cases}$$

Cálculo das probabilidades

- (a) traçar o gráfico de F .
 - (b) Descrever a função massa probabilística p e traçar o respectivo gráfico.
 - (c) Calcular as seguintes probabilidades: $P(X = 1)$, $P(X \leq 1)$, $P(X < 1)$, $P(X = 2)$, $P(X \leq 2)$, $P(0 < X < 2)$, $P(0 < X \leq 2)$, $P(1 \leq X \leq 2)$.
4. Considerar uma variável aleatória X cujos valores possíveis são todos os números racionais da forma $\frac{n}{n+1}$ e $\frac{n+1}{n}$, com $n = 1, 2, 3, \dots$. Se

$$P\left(X = \frac{n}{n+1}\right) = P\left(X = \frac{n+1}{n}\right) = \frac{1}{2^{n+1}},$$

- verificar que é possível esta definição de probabilidades e esboçar a forma geral do gráfico da função de repartição F_X .
5. A função massa probabilística p de uma variável aleatória X é zero excepto nos pontos $t = 0, 1, 2$. Nesses pontos toma os valores

$$p(0) = 3c^3, \quad p(1) = 4c - 10c^2, \quad p(2) = 5c - 1,$$

para algum $c > 0$.

- (a) Determinar o valor de c .
 - (b) Calcular as seguintes probabilidades: $P(X < 1)$, $P(X < 2)$, $P(1 < X \leq 2)$, $P(0 < X < 3)$.
 - (c) Definir a função de repartição F e traçar o seu gráfico.
 - (d) Determinar o maior valor de t tal que $F(t) < \frac{1}{2}$.
 - (e) Determinar o menor valor de t tal que $F(t) > \frac{1}{2}$.
6. Uma variável aleatória X tem uma distribuição binomial com parâmetros $n = 4$ e $p = \frac{1}{3}$.
- (a) Definir a função massa probabilística p , e traçar o seu gráfico.
 - (b) Definir a função de repartição F , e traçar o seu gráfico.
 - (c) Calcular as probabilidades $P(1 < X \leq 2)$ e $P(1 \leq X \leq 2)$.
7. Supor que se lança sobre uma mesa uma brocha, a qual cai ou com a ponta para cima ou numa posição estável com a ponta sobre a mesa. Supor que existe uma probabilidade p de que fique com a ponta para cima:
- (a) Admitindo que se lançam sobre a mesa duas brochas idênticas, e supondo que os acontecimentos são aleatoriamente independentes, provar que a probabilidade de que ambas fiquem com a ponta para cima é p^2 .
 - (b) Continuando a alínea (a), seja X a variável aleatória definida pelo número de brochas que ficam com a ponta para cima (os valores possíveis de X são 0, 1 e 2). Calcular as probabilidades $P(X = 0)$ e $P(X = 1)$.
 - (c) Representar a função de repartição F_X quando $p = \frac{1}{3}$.
8. Dada uma variável aleatória X cujos valores possíveis são 1, 2, ..., n , supor que a probabilidade $P(X = k)$ é proporcional a k . Determinar a constante de proporcionalidade, a função massa probabilística p_X , e a função de repartição F_X .
9. Dada a variável aleatória X cujos valores possíveis são 0, 1, 2, 3, ..., supor que $P(X = k)$ é proporcional a $c^k/k!$ onde c é um número real fixo. Determinar a constante de proporcionalidade e a função massa probabilística p .

10. (a) Lança-se um dado perfeito. O conjunto fundamental é $S = \{1, 2, 3, 4, 5, 6\}$. Se o número de pontos obtidos for ímpar um jogador recebe um escudo; caso contrário pagará um escudo. Seja X a variável aleatória que define os ganhos do jogador em cada jogada (número de escudos). (Os valores possíveis de X são $+1$ e -1 .) Definir a função massa probabilística p_X e a função de repartição F_X . Traçar os respectivos gráficos.
- (b) Lança-se uma moeda perfeita ao ar. O conjunto fundamental é $S = \{H, T\}$. Se o resultado é a face "cara" o jogador recebe um escudo; se sai a face "cruz" o jogador paga um escudo. Seja Y a variável aleatória que define o ganho do jogador (número de escudos) em cada jogada. Provar que a função massa probabilística p_Y e a função de repartição F_Y são idênticas às da alínea (a). Mostra este exemplo que diferentes variáveis aleatórias podem ter a mesma função de repartição. Existem infinitas variáveis aleatórias tendo uma dada função de repartição F . (Porquê?) Tais variáveis aleatórias dizem-se estarem *identicamente distribuídas*. Todo o teorema relativo a uma função de repartição particular é aplicável a qualquer das infinitas variáveis aleatórias possuindo essa função de repartição.
11. Sabe-se que o número de minutos que uma pessoa tem que esperar por um comboio em certa estação do metropolitano é uma variável aleatória X com a seguinte função massa probabilística:

$$p(t) = 0 \text{ excepto para } t = 3k/10 \quad \text{se } k = 0, 1, 2, \dots, 10.$$

$$p(t) = \frac{1}{12} \quad \text{se } t = 0, 0.3, 0.6, 0.9, 2.1, 2.4, 2.7, 3.0.$$

$$p(t) = \frac{1}{6} \quad \text{se } t = 1.2, 1.5, 1.8.$$

Traçar o gráfico da correspondente função de repartição F . Seja A o acontecimento que consiste em uma pessoa ter que esperar entre 0 e 2 minutos (incluindo 0 e 2), e seja B o acontecimento de ter que esperar entre 1 e 3 minutos (incluindo 1 e 3). Calcular as probabilidades: $P(A)$, $P(B)$, $P(A \cap B)$, $P(B \cup A)$, $P(A \cup B)$.

12. (a) Se $0 < p < 1$ e $q = 1 - p$, mostrar que

$$\binom{n}{k} p^k q^{n-k} = \frac{(np)^k}{k!} \left(1 - \frac{np}{n}\right)^n Q_n,$$

sendo

$$Q_n = \frac{\prod_{r=2}^k \left(1 - \frac{r-1}{n}\right)}{(1-p)^k}.$$

- (b) Dado $\lambda > 0$, seja $p = \lambda/n$ para $n > \lambda$. Demonstrar que $Q_n \rightarrow 1$ quando $n \rightarrow \infty$ e que

$$\binom{n}{k} p^k q^{n-k} \rightarrow \frac{\lambda^k}{k!} e^{-\lambda} \text{ quando } n \rightarrow \infty.$$

Este resultado sugere que, para grandes valores de n e pequenos valores de p , a distribuição binomial é aproximadamente a mesma que a distribuição de Poisson, com tanto que o produto np seja aproximadamente constante; este valor constante é o parâmetro λ da distribuição de Poisson.

14.9. Distribuições contínuas. Funções densidade

Seja X uma variável aleatória unidimensional e F a sua função de repartição, de modo que $F(t) = P(X \leq t)$ para todo o real t . Se a probabilidade $P(X = t)$ é zero para todo t então, pelo teorema 14.5, F é contínua em todo o eixo real. Neste caso F diz-se uma *distribuição contínua* e X chama-se uma *variável aleatória contínua*. Se a derivada F' existe e é contínua num intervalo $[a, t]$ o segundo teorema fundamental do cálculo permite-nos escrever

$$(14.13) \quad F(t) - F(a) = \int_a^t f(u) du,$$

sendo f a derivada de F . A diferença $F(t) - F(a)$ é, naturalmente, a probabilidade $P(a < x \leq t)$, e a equação (14.13) exprime esta probabilidade por intermédio de um integral.

Algumas vezes a função de repartição F pode exprimir-se por um integral da forma (14.13), no qual o integrando f é integrável mas não necessariamente contínuo. Sempre que uma equação tal como (14.13) seja válida para todos os intervalos $[a, t]$, o integrando f chama-se a *função densidade de probabilidade* da variável aleatória X (ou da função de repartição F), desde que f seja não negativa. Por outras palavras, temos a seguinte definição:

DEFINIÇÃO DA FUNÇÃO DENSIDADE DE PROBABILIDADE. *Seja X uma variável aleatória unidimensional com uma função de repartição F , contínua. Uma função não negativa f diz-se uma função densidade de probabilidade de X (ou de F) se for integrável em todo o intervalo $[a, t]$ e se*

$$(14.14) \quad F(t) - F(a) = \int_a^t f(u) du.$$

Se em (14.14) fizermos $a \rightarrow -\infty$, então $F(a) \rightarrow 0$ e obtemos a importante fórmula

$$(14.15) \quad F(t) = P(X \leq t) = \int_{-\infty}^t f(u) du,$$

válida para todo o real t . Se fizermos agora $t \rightarrow +\infty$, e tivermos presente que $F(t) \rightarrow 1$, obtemos

$$(14.16) \quad \int_{-\infty}^{+\infty} f(u) du = 1.$$

Para variáveis aleatórias discretas a soma de todas as probabilidades $P(X = t)$ é igual a 1. A fórmula (14.16) é a proposição análoga para o caso contínuo. Existe igualmente uma forte analogia entre as fórmulas (14.11) e (14.15). A função densidade desempenha o mesmo papel para as distribuições contínuas que a função massa probabilística p para as distribuições discretas—a integração substitui a soma no cálculo

das probabilidades. Há, contudo, uma importante diferença. No caso discreto $p(t)$ é a probabilidade de que $X = t$, mas no caso contínuo $f(t)$ não é a probabilidade de que $X = t$. Com efeito, esta probabilidade é zero porque F é contínua para todo t . Naturalmente, isto significa também que para uma distribuição contínua se tem

$$P(a \leq X \leq b) = P(a < X < b) = P(a < X \leq b) = P(a \leq X < b).$$

Se F tem uma densidade f cada uma dessas probabilidades é igual à integral $\int_a^b f(u) du$.

Nota: Uma dada distribuição pode admitir mais do que uma função densidade, visto que o valor do integrando em (14.14) pode mudar-se num número finito de pontos sem que por isso se altere o integral. Mas se f é contínua em t então $f(t) = F'(t)$; neste caso o valor da função densidade em t fica completamente determinado por F .

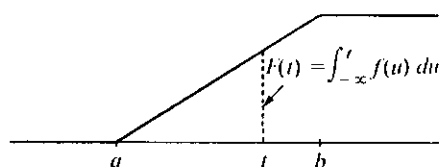
Visto f ser não negativa, o segundo membro de (14.14) pode interpretar-se geometricamente como definindo a área da porção do conjunto de ordenadas de f situado à esquerda da recta $x = t$. A área de todo o conjunto de ordenadas é igual a 1. A área da parte do conjunto de ordenadas correspondente a um dado intervalo (seja aberto, fechado ou semi-aberto) define a probabilidade de que a variável aleatória X assuma um valor naquele intervalo. A figura 14.6 mostra um exemplo de uma função de repartição contínua F e a correspondente função densidade f . A ordenada $F(t)$, na figura 14.6 (a), é igual à área da parte sombreada da figura 14.6 (b).

Nas Secções seguintes vamos estudar alguns exemplos importantes de distribuições contínuas.

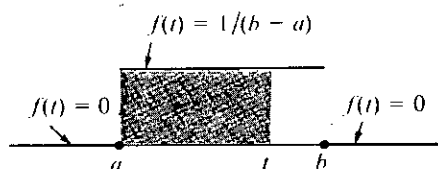
14.10. Distribuição uniforme num intervalo

Uma variável aleatória X , unidimensional, diz-se admitir uma função de repartição uniforme F sobre um intervalo finito $[a, b]$ se F for definida do modo seguinte:

$$F(t) = \begin{cases} 0 & \text{se } t \leq a, \\ \frac{t-a}{b-a} & \text{se } a < t < b, \\ 1 & \text{se } t \geq b. \end{cases}$$



(a) A função de repartição F .



(b) A função densidade f .

FIG. 14.6. Uma distribuição uniforme sobre um intervalo $[a, b]$ e a correspondente função densidade.

Esta é uma distribuição contínua cujo gráfico está representado na figura 14.6 (a).

A derivada $F'(t)$ existe para todo o valor de t , excepto nos pontos a e b e podemos escrever

$$F(t) = \int_{-\infty}^t f(u) du,$$

onde f é a função densidade definida do modo seguinte:

$$f(t) = \begin{cases} 1/(b-a) & \text{se } a < t < b, \\ 0 & \text{para os restantes valores } t. \end{cases}$$

O gráfico de f está traçado na figura 14.6(b).

O teorema que apresentamos a seguir caracteriza distribuições uniformes de outro modo.

TEOREMA 14.7. *Sejam X uma variável aleatória unidimensional com todos os seus valores num intervalo finito $[a, b]$, e F a função de repartição de X . A função F é uniforme em $[a, b]$ se e só se*

$$(14.17) \quad P(X \in I) = P(X \in J)$$

para todo o par de subintervalos I e J de $[a, b]$ com a mesma amplitude, caso em que se tem

$$P(X \in I) = \frac{h}{b-a},$$

com h a amplitude de I .

Demonstração. Admitamos primeiramente que X tem uma distribuição uniforme sobre $[a, b]$. Se $[c, c+h]$ é qualquer subintervalo de $[a, b]$ de amplitude h temos

$$P(c \leq X \leq c+h) = F(c+h) - F(c) = \frac{c+h-a}{b-a} - \frac{c-a}{b-a} = \frac{h}{b-a}.$$

Isto prova que $P(X \in I) = P(X \in J) = h/(b-a)$, para todo o par de subintervalos I e J de $[a, b]$ de amplitude h .

Para demonstrar o recíproco, suponhamos que X satisfaz a (14.17). Observemos em primeiro lugar que $F(t) = 0$ se $t < a$ e $F(t) = 1$ se $t > b$, uma vez que X tem todos os seus valores em $[a, b]$.

Introduzamos uma nova função g definida no intervalo $(0, b-a]$ por

$$(14.18) \quad g(u) = P(a < X \leq a+u) \quad \text{se } 0 < u \leq b-a.$$

Recorrendo à aditividade e à propriedade (14.17) encontramos

$$\begin{aligned}
 g(u+v) &= P(a < X \leq a+u+v) \\
 &= P(a < X \leq a+u) + P(a+u < X \leq a+u+v) \\
 &= g(u) + P(a < X \leq a+v) = g(u) + g(v),
 \end{aligned}$$

desde que $0 < u+v \leq b-a$, isto é, g satisfaz à equação funcional

$$g(u+v) = g(u) + g(v)$$

para todas as funções u e v tais que $u > 0$, $v > 0$, $u+v \leq b-a$. Esta é a chamada *equação funcional de Cauchy*. Demonstraremos um pouco mais adiante que toda a solução não negativa da equação funcional de Cauchy é dada por

$$g(u) = \frac{u}{b-a} g(b-a) \quad \text{para } 0 < u \leq b-a.$$

Aplicando isto na equação (14.18) encontramos que para $0 < u \leq b-a$ se tem

$$P(a < X \leq a+u) = \frac{u}{b-a} P(a < X \leq b) = \frac{u}{b-a}$$

uma vez que $P(a < X \leq b) = 1$. Por outras palavras,

$$F(a+u) - F(a) = \frac{u}{b-a} \quad \text{se } 0 < u \leq b-a.$$

Fazendo $t = a+u$, podemos escrever a igualdade anterior na forma

$$F(t) - F(a) = \frac{t-a}{b-a} \quad \text{se } a < t \leq b.$$

Mas $F(a) = 0$, uma vez que F é contínua à direita. Logo

$$F(t) = \frac{t-a}{b-a} \quad \text{se } a \leq t \leq b,$$

o que prova que F é uniforme em $[a, b]$.

TEOREMA 14.8. SOLUÇÃO DA EQUAÇÃO FUNCIONAL DE CAUCHY. *Seja g uma função real definida no intervalo semi-aberto $(0, c]$ e gozando das duas propriedades seguintes:*

- (a) $g(u+v) = g(u) + g(v)$ sempre que u, v e $u+v$ pertencem a $(0, c]$.
- (b) g é não negativa em $(0, c]$.

Então g vem dada pela fórmula

$$g(u) = \frac{u}{c} g(c) \quad \text{para } 0 < u \leq c.$$

Demonstração. Por uma mudança de escala podemos reduzir a demonstração ao caso particular em que $c = 1$. Com efeito, seja

$$G(x) = g(cx) \quad \text{para } 0 < x \leq 1.$$

Então G é não negativa e satisfaz à equação funcional de Cauchy

$$G(x + y) = G(x) + G(y)$$

sempre que x, y e $x + y$ pertencem a $(0, 1]$. Se provarmos que

$$(14.19) \quad G(x) = xG(1) \quad \text{para } 0 < x \leq 1$$

resulta que $g(cx) = xg(c)$, ou que $g(u) = (u/c)g(c)$ para $0 < u \leq c$.

Se x pertence ao intervalo $(0, 1]$ então $x/2$ pertence igualmente ao mesmo intervalo e tem-se

$$G(x) = G\left(\frac{x}{2}\right) + G\left(\frac{x}{2}\right) = 2G\left(\frac{x}{2}\right).$$

Por indução, para cada x em $(0, 1]$ tem-se

$$(14.20) \quad G(x) = nG\left(\frac{x}{n}\right) \quad \text{para } n = 1, 2, 3, \dots$$

Analogamente, se y e my pertencem a $(0, 1]$ tem-se

$$G(my) = mG(y) \quad \text{para } m = 1, 2, 3, \dots$$

Fazendo $y = x/n$ e utilizando (14.20) obtemos

$$G\left(\frac{m}{n}x\right) = \frac{m}{n}G(x)$$

se x e mx/n pertencem a $(0, 1]$. Por outras palavras, temos

$$(14.21) \quad G(rx) = rG(x)$$

para todo o número racional positivo r tal que x e rx pertencem a $(0, 1]$.

Consideremos agora um x qualquer no intervalo aberto $(0, 1)$ e sejam $\{r_n\}$ e $\{R_n\}$ duas sucessões racionais em $(0, 1]$ tais que

$$r_n < x < R_n \quad \text{e} \quad \lim_{n \rightarrow \infty} r_n = \lim_{n \rightarrow \infty} R_n = x.$$

A equação funcional de Cauchy e a propriedade de G ser não negativa mostram que $G(x + y) \geq G(x)$, pelo que G é monótona crescente em $(0, 1]$. Por conseguinte

$$G(r_n) \leq G(x) \leq G(R_n).$$

Aplicando (14.21) podemos escrever de novo

$$r_n G(1) \leq G(x) \leq R_n G(1).$$

Fazendo $n \rightarrow \infty$ encontramos $xG(1) \leq G(x) \leq xG(1)$, pelo que $G(x) = xG(1)$, o que prova (14.19).

Nota: As distribuições uniformes são frequentes em experiências cujos resultados são pontos escolhidos ao acaso num intervalo $[a, b]$, ou em experiências que se referem a um intervalo $[a, b]$ como um alvo onde é impossível fazer pontaria. Os termos “ao acaso” e “impossível fazer pontaria” são usualmente interpretados como significando que se I é qualquer subintervalo de $[a, b]$, então a probabilidade $P(X \in I)$ depende unicamente da amplitude de I e não da sua localização em $[a, b]$. O teorema 14.7 mostra que as distribuições uniformes são as únicas distribuições com estas propriedades.

Voltemos agora a nossa atenção para as questões de probabilidades formuladas no início do capítulo.

EXEMPLO. Divide-se um segmento de recta em duas partes, determinando ao acaso o ponto de divisão. Seja X a variável aleatória definida pelo cociente do comprimento da parte esquerda pelo comprimento da parte direita. Determinar a função de repartição F_X .

Resolução. Usemos o intervalo $[0, 1]$ para representar o segmento de recta e exprimamos o ponto de divisão pela variável aleatória $Y(\omega) = \omega$ para cada ω em $(0, 1)$. Uma vez que o ponto de divisão é escolhido ao acaso admitimos que a Y corresponde uma função de repartição uniforme F_Y sobre $[0, 1]$. Logo

$$F_Y(t) = t \quad \text{para } 0 \leq t \leq 1.$$

Se o segmento se divide pelo ponto ω , então $\omega/(1 - \omega)$ é o cociente do comprimento da parte esquerda pelo da parte direita. Deste modo $X(\omega) = \omega/(1 - \omega)$.

Se $t < 0$ temos $F_X(t) = 0$, uma vez que o cociente $X(\omega)$ não pode ser negativo. Se $t \geq 0$ a desigualdade $X(\omega) \leq t$ é equivalente a $\omega/(1 - \omega) \leq t$, a qual é equivalente a $\omega \leq t/(1 + t)$. Portanto

$$F_X(t) = P(X \leq t) = P\left(Y \leq \frac{t}{1+t}\right) = F_Y\left(\frac{t}{1+t}\right) = \frac{t}{1+t}$$

visto que $0 \leq t/(1 + t) < 1$.

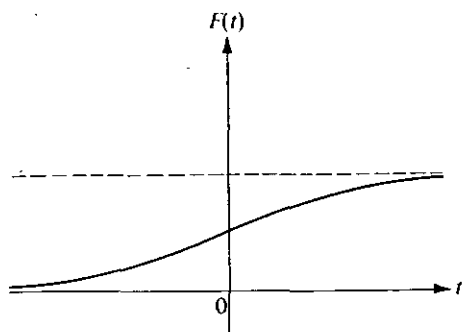
Podemos agora calcular diversas probabilidades. Por exemplo, a probabilidade de que as duas partes tenham igual comprimento é $P(X = 1) = 0$. Com efeito, porque F_X é uma função de repartição contínua, a probabilidade de que X tome um valor particular é zero.

A probabilidade de que o segmento da esquerda tenha pelo menos o dobro do comprimento do segmento da direita é $P(X \geq 2) = 1 - P(X < 2) = 1 - \frac{2}{3} = \frac{1}{3}$. Analogamente, a probabilidade de que o segmento da direita tenha pelo menos o dobro do comprimento do segmento da esquerda é $P(X \leq \frac{1}{2}) = \frac{1}{3}$. A probabilidade de que o segmento mais comprido tenha pelo menos o dobro do comprimento do segmento mais curto é $P(X \geq 2) + P(X \leq \frac{1}{2}) = \frac{2}{3}$.

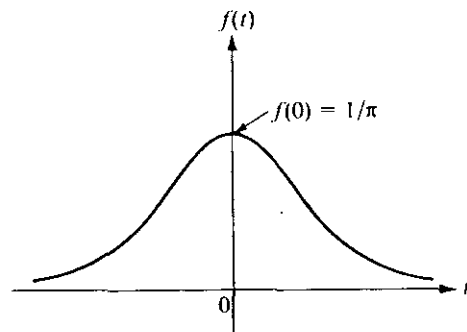
14.11. Distribuição de Cauchy

Uma variável aleatória X diz-se uma distribuição de Cauchy F se

$$F(t) = \frac{1}{2} + \frac{1}{\pi} \arctan t$$



(a) A função de repartição F



(b) A função densidade f .

FIG. 14.7 Função de repartição de Cauchy e correspondente função densidade.

para todo o real t . Esta função admite derivada contínua para todo o valor de t ; a correspondente função densidade, contínua, é dada pela fórmula

$$f(t) = \frac{1}{\pi(1 + t^2)}.$$

Os gráficos de F e f estão representados nas figuras 14.7(a) e (b) respectivamente.

A experiência que vamos referir conduz a uma distribuição de Cauchy. Uma agulha que pode girar em torno do ponto $(-1, 0)$ é posta em movimento e deixa-se parar ao acaso. Um resultado da experiência é θ , o ângulo da agulha com o eixo OX , o qual se mede de modo que $-\pi/2 < \theta \leq \pi/2$. Seja X a variável aleatória definida por $X(\theta) = \theta$

que admite uma distribuição uniforme no intervalo $[-\pi/2, \pi/2]$ com densidade $f_X(t) = 1/\pi$ se $-\pi/2 < t < \pi/2$.

Consideremos agora a variável aleatória Y definida por.

$$Y(\theta) = \tan \theta.$$

Vamos provar que Y admite uma distribuição de Cauchy F_Y se X admitir uma distribuição uniforme em $[-\pi/2, \pi/2]$.

Se $a < t$ seja $\alpha = \arctg a$ e $\theta = \arctg t$. Então tem-se

$$F_Y(t) - F_Y(a) = P(a < Y \leq t) = P(\alpha < X \leq \theta) = \int_{\alpha}^{\theta} f_X(u) du = \frac{\theta - \alpha}{\pi}.$$

Porque $\alpha \rightarrow -\pi/2$ quando $a \rightarrow -\infty$ obtemos

$$F_Y(t) = \frac{\theta + \frac{1}{2}\pi}{\pi} = \frac{1}{\pi} \arctan t + \frac{1}{2}.$$

Isto prova que Y admite uma distribuição de Cauchy.

14.12. Exercícios

1. Uma variável aleatória X tem uma função de repartição contínua F , definida do modo seguinte

$$F(t) = \begin{cases} 0 & \text{se } t \leq 0, \\ ct & \text{se } 0 \leq t \leq 1, \\ 1 & \text{se } t > 1. \end{cases}$$

- (a) Determinar a constante c e definir a função densidade f .
 - (b) Calcular as probabilidades $P(X = \frac{1}{3})$, $P(X < \frac{1}{3})$, $P(|X| < \frac{1}{3})$.
2. Seja $f(t) = c|\sin t|$ para $|t| < \pi/2$ e $f(t) = 0$ para os restantes valores de t . Determinar o valor da constante c de modo que f seja a função densidade de uma função de repartição contínua F . Determinar também F e traçar o respectivo gráfico.
 3. Resolver o Exercício 2 se $f(t) = c(4t - 2t^2)$ para $0 \leq t \leq 2$, e $f(t) = 0$ para os restantes valores de t .
 4. O tempo em minutos que uma pessoa é obrigada a esperar por um autocarro é uma variável aleatória com a função densidade f definida por

$$f(t) = \frac{1}{2} \quad \text{para } 0 < t < 1, \quad f(t) = \frac{1}{4} \quad \text{para } 2 < t < 4, \quad f(t) = 0 \quad \text{para os restantes valores de } t.$$

Calcular a probabilidade de que o tempo de espera dessa pessoa seja

- (a) superior a um minuto;
- (b) superior a dois minutos;
- (c) superior a três minutos.

5. Uma variável aleatória X tem uma função de repartição contínua F e a função densidade f . A função densidade goza das seguintes propriedades: $f(t) = 0$ se $t < \frac{1}{2}$, $f(\frac{1}{2}) = 1$, $f(t)$ é linear se $\frac{1}{2} \leq t \leq 1$, $f(1 - t) = f(t)$ para todo t .
- Traçar o gráfico de f .
 - Dar um conjunto de fórmulas que definam F e traçar o seu gráfico.
 - Calcular as probabilidades seguintes: $P(X < 1)$, $P(X < \frac{1}{2})$, $P(X < \frac{1}{2})$, $P(X \leq \frac{1}{2})$, $P(\frac{1}{2} < X < \frac{3}{4})$.
6. Uma variável aleatória X tem uma distribuição uniforme em $[-3, 3]$.
- Calcular $P(X = 2)$, $P(X < 2)$, $P(|X| < 2)$, $P(|X - 2| < 2)$.
 - Determinar t para o qual $P(X > t) = \frac{1}{3}$.
7. A "Lethe Subway Company" tem em determinada estação um comboio, na direcção norte, cada 30 minutos. Um homem entra na estação num instante ao acaso. Seja X a variável aleatória definida pelo número de minutos que o homem tem que esperar até chegar o primeiro comboio. Suponhamos que X tem uma distribuição uniforme no intervalo $[0, 30]$. (É este o significado a atribuir à expressão "entra na estação num instante ao acaso").
- Para cada $k = 5, 10, 15, 20, 25, 30$, calcular a probabilidade de que ele tenha de esperar k minutos pelo primeiro comboio e chegar à estação.
 - Uma empresa concorrente, a Styx Subway Company, está autorizada a fazer passar na mesma estação e em direcção ao norte, um comboio cada 30 minutos, mas devendo decorrer pelo menos 5 minutos entre as chegadas à estação de dois comboios de empresa diferentes. Suponhamos que os passageiros entram na estação em instantes ao acaso e tomam sempre o primeiro comboio a chegar à estação. Provar que a Styx Company pode organizar o seu horário de modo a que transporte cinco vezes mais passageiros que a outra empresa.
8. Seja X uma variável aleatória com a função de repartição F_X no intervalo $[0, 1]$. Seja $Y = aX + b$, com $a > 0$. Determinar a função de repartição F_Y e traçar o respectivo gráfico.
9. Uma roda de roleta contém os inteiros de 0 a 36, distribuídos por 37 arcos iguais. Põe-se a roda a girar e deixa-se parar ao acaso, registando-se o número da roda mais próximo de uma marca fixa. Considere-se este ponto de paragem como uma variável aleatória X com uma distribuição uniforme. Calcular a probabilidade de que X esteja num arco contendo (a) o inteiro 0; (b) um inteiro n no intervalo $11 \leq n \leq 20$; (c) um inteiro ímpar.
10. Diz-se que uma variável aleatória tem uma distribuição de Cauchy com parâmetros a e b , com $a > 0$, se a sua função densidade é definida por

$$f(t) = \frac{1}{\pi} \frac{a}{a^2 + (t - b)^2}.$$

Verificar que o integral de f de $-\infty$ a $+\infty$ é 1, e determinar a função de repartição F .

11. Seja $f_1(t) = 1$ para $0 < t < 1$, e seja $f_1(t) = 0$ para os restantes valores de t . Defina-se uma sucessão de funções $\{f_n\}$ pela fórmula

$$f_{n+1}(x) = \int_{-\infty}^{\infty} f_1(x - t)f_n(t)dt.$$

- Provar que $f_{n+1}(x) = \int_{x-1}^x f_n(t)dt$
- Traçar os gráficos de f_1, f_2 e f_3 .

12. Considerar o Exercício 11. Provar que cada função f_n é uma densidade de probabilidade.

14.13. Distribuições exponenciais

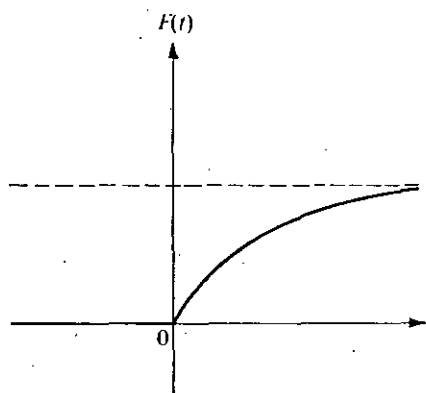
Seja λ uma constante positiva. Uma variável aleatória unidimensional X diz-se ter uma distribuição exponencial F de parâmetro λ se

$$F(t) = \begin{cases} 1 - e^{-\lambda t} & \text{para } t \geq 0, \\ 0 & \text{para } t < 0. \end{cases}$$

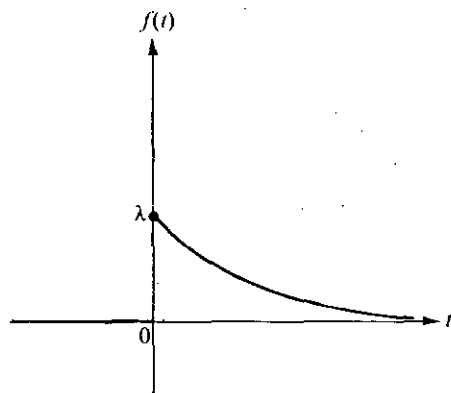
A correspondente função densidade f é dada pela fórmula

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & \text{para } t \geq 0, \\ 0 & \text{para } t < 0. \end{cases}$$

Os gráficos de F e f são semelhantes aos representados na figura 14.8.



(a) A função de repartição F .



(b) A função densidade f .

FIG. 14.8 Uma distribuição exponencial e a correspondente função densidade.

As distribuições exponenciais gozam de uma propriedade característica que sugere a sua aplicação em certos problemas relativos a desintegração radioactiva, acidentes de tráfego, e avarias em equipamentos electrónicos tais como válvulas electrónicas. Esta propriedade é análoga aquela que caracteriza as distribuições uniformes e pode descrever-se do modo seguinte:

Designamos por X o tempo decorrido até que uma peça do equipamento avarie e por F a função de repartição de X . Admitamos que $F(t) = 0$ para $t \leq 0$, não se impondo de momento outras restrições a F . Se $t > 0$, então $X \leq t$ é o acontecimento “ocorre a avaria no intervalo de tempo $[0, t]$ ”. Logo $X > t$ é o acontecimento complementar, “não ocorre avaria no intervalo de tempo $[0, t]$ ”.

Suponhamos que não ocorre qualquer avaria no intervalo de tempo $[0, t]$. Qual é a probabilidade de que o mesmo se verifique no intervalo de tempo $[t, t + s]$? Este é um problema de probabilidades condicionadas. Pretendemos determinar $P(X > t + s | X > t)$, a probabilidade de que não ocorre qualquer avaria no intervalo de tempo $[0, t + s]$, sabendo que ela não ocorre de certeza no intervalo de tempo $[0, t]$.

Segundo a definição de probabilidade condicional temos

$$(14.22) \quad P(X > t + s \mid X > t) = \frac{P[(X > t + s) \cap (X > t)]}{P(X > t)} = \frac{P(X > t + s)}{P(X > t)}.$$

Suponhamos agora que F é uma distribuição exponencial de parâmetro $\lambda > 0$. Então $F(t) = 1 - e^{-\lambda t}$ para $t > 0$, e $P(X > t) = 1 - P(X \leq t) = e^{-\lambda t}$. Logo (14.22) escreve-se

$$P(X > t + s \mid X > t) = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} = P(X > s).$$

Por outras palavras, se a peça do equipamento sobrevive no intervalo de tempo $[0, t]$, então a probabilidade de assim continuar no intervalo $[t, t + s]$ é igual à probabilidade de sobrevivência no intervalo $[0, s]$ com a mesma amplitude. Quer isto dizer que a probabilidade de sobrevivência depende unicamente da amplitude do intervalo de tempo e não da idade do equipamento. Expressa em termos da função de repartição F , esta propriedade estabelece que

$$(14.23) \quad \frac{1 - F(t + s)}{1 - F(t)} = 1 - F(s) \quad \text{para todo } t > 0 \text{ e } s > 0.$$

O teorema seguinte mostra que as distribuições exponenciais são as únicas gozando desta propriedade.

TEOREMA 14.9. *Se F é uma função de repartição de probabilidade satisfazendo à equação funcional (14.23), com $F(t) < 1$ para $t > 0$, então existe uma constante positiva $\lambda > 0$ tal que*

$$F(t) = 1 - e^{-\lambda t} \quad \text{para todo } t > 0.$$

Demonstração. Seja $g(t) = -\log [1 - F(t)]$ para $t > 0$. Então $1 - F(t) = e^{-g(t)}$, de modo que para demonstrar o teorema basta demonstrar que $g(t) = \lambda t$ para algum $\lambda > 0$.

Agora g é não negativa e satisfaz à equação funcional de Cauchy

$$g(t + s) = g(t) + g(s)$$

para todo $t > 0$ e $s > 0$. Deste modo, aplicando o teorema 14.8 com $c = 1$, deduzimos que $g(t) = tg(1)$ para $0 < t \leq 1$. Seja $\lambda = g(1)$. Então $\lambda = -\log [1 - F(1)] > 0$ e por conseguinte $g(t) = \lambda t$ para $0 < t \leq 1$.

Para demonstrar que $g(t) = \lambda t$ para todo $t > 0$, façamos $G(t) = g(t) - \lambda t$. A função G satisfaz também à equação funcional de Cauchy. Além disso, G é periódica de período 1 porque $G(t + 1) = G(t) + G(1)$ e $G(1) = 0$. Uma vez que G é identicamente nula

em $[0, 1]$ a periodicidade mostra que $G(t) = 0$ para todo $t > 0$. Por outras palavras, $g(t) = \lambda t$ para todo $t > 0$, o que completa a demonstração.

EXEMPLO 1. Seja X uma variável aleatória que mede o tempo de vida (em horas) de um certo tipo de válvula electrónica. Admitamos que X tem uma distribuição exponencial com parâmetro $\lambda = 0,001$. O fabricante deseja garantir as válvulas para uma duração de T horas. Determinar T de modo que $P(X > T) = 0,95$.

Resolução. A função de repartição é dada por $F(t) = 1 - e^{-\lambda t}$ para $t > 0$, com $\lambda = 0,001$. Uma vez que $P(X > T) = 1 - F(T) = e^{-\lambda T}$, escolhamos T de modo que $e^{-\lambda T} = 0,95$. Logo $T = -(\log 0,95)/\lambda = -1000 \log 0,95 = 51,25+$.

EXEMPLO 2. Consideremos a variável aleatória do Exemplo 1, mas com o valor de λ não especificado. O raciocínio seguinte sugere um processo razoável de determinação de λ . Partimos com um número inicial de válvulas electrónicas no instante $t = 0$, e representamos por $g(t)$ o número de válvulas ainda a funcionarem ao fim de t horas. O cociente $[g(0) - g(t)]/g(0)$ é a fracção do número original que avariaram no tempo t . Porque a probabilidade de que certa válvula avarie no tempo t é $1 - e^{-\lambda t}$, parece razoável esperar que a equação

$$(14.24) \quad \frac{g(0) - g(t)}{g(0)} = 1 - e^{-\lambda t}$$

seja uma boa aproximação da realidade. Se aceitamos (14.24), obtemos

$$g(t) = g(0)e^{-\lambda t}.$$

Por outras palavras, sob as hipóteses (14.24), o número $g(t)$ obedece a uma lei exponencial de deterioração com a constante de deterioração λ . A constante λ pode calcular-se em termos da vida média. Se t_1 é a vida média então $\frac{1}{2} = g(t_1)/g(0) = e^{-\lambda t_1}$, pelo que $\lambda = (\log 2)/t_1$. Por exemplo, se a vida média de uma grande amostra de válvulas é conhecida e igual a 693 horas, obtemos $\lambda = (\log 2)/693 = 0,001$.

14.14. Distribuições normais

Sejam m e σ números reais fixos, com $\sigma > 0$. Uma variável aleatória X diz-se ter uma *distribuição normal* de média m e variância σ^2 se a função densidade f é definida pela fórmula

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-[(t-m)/\sigma]^2/2}$$

para todo o real t . A correspondente função de repartição F é, naturalmente, o integral

$$F(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^t e^{-[(u-m)/\sigma]^2/2} du.$$

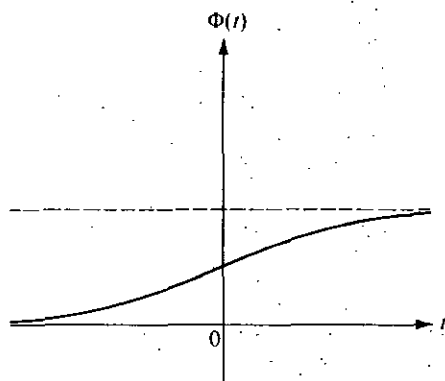


FIG. 14.9 Função de repartição da variável normal reduzida (ou típica): $m = 0$, $\sigma = 1$.

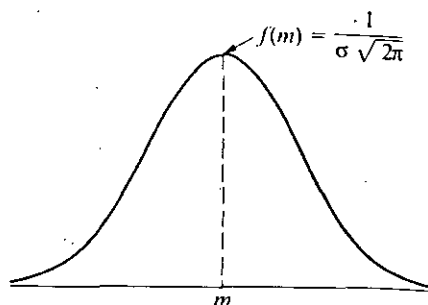


FIG. 14.10 A função densidade de uma distribuição normal com média m e variância σ^2 .

É evidente que esta função F é monótona crescente, contínua para todo o valor de t e tende para 0 quando $t \rightarrow -\infty$. Também se pode provar que $F(t) \rightarrow 1$ quando $t \rightarrow +\infty$. (Ver Exercício 7 da Secção 14.16).

O caso particular em que $m = 0$, $\sigma = 1$ chama-se a distribuição normal *standard* (ou *padrão*). Neste caso a função de repartição F representa-se habitualmente pela letra Φ . Assim,

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-u^2/2} du.$$

O caso geral pode reduzir-se ao caso *standard* pela mudança de variável $v = (u - m)/\sigma$ no integral de F . Isto conduz-nos à fórmula

$$F(t) = \Phi\left(\frac{t - m}{\sigma}\right).$$

A tábuia 14.1 é uma tábuia de valores de $\Phi(t)$ com quatro casos decimais para valores de t espaçados em intervalos de amplitude 0,01 desde $t = 0,000$ até $t = 3,69$. Na figura 14.9 está representado o gráfico de $\Phi(t)$. O gráfico da função densidade f é a famosa curva em “forma de sino” desenhada na figura 14.10. O máximo da curva corresponde à média m . Para grandes valores de σ a curva achata-se, e para σ pequeno tem a forma de um cume como na figura 14.10.

As distribuições normais estão entre os mais importantes de todas as distribuições contínuas. A muitas variáveis aleatórias que ocorrem na natureza correspondem funções de repartição normais, ou aproximadamente normais. Como exemplos citaremos a medida da estatura dos indivíduos de uma grande população, certas medidas sobre grandes populações de seres vivos encontrados na Biologia e os erros de observação cometidos ao fazer um grande número de medidas. Na Física, a lei de Maxwell das

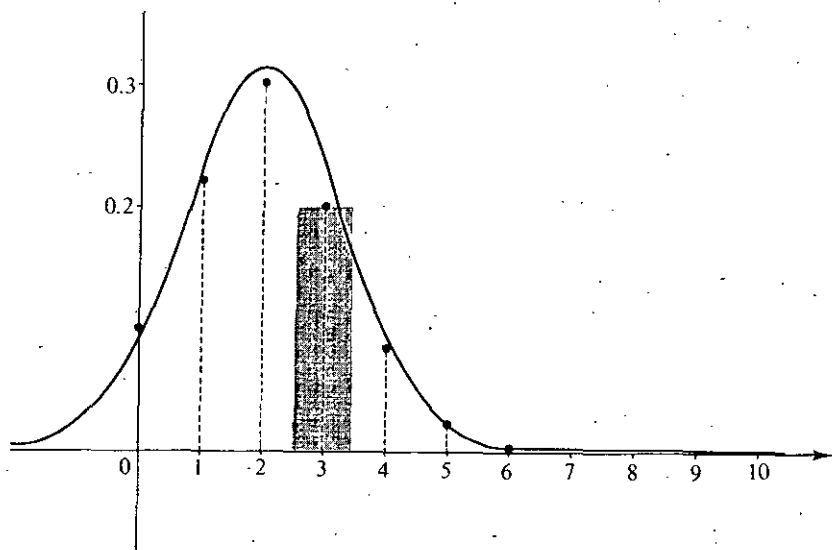


FIG. 14.11 A função densidade de uma distribuição normal considerada com uma aproximação da função massa probabilística de uma distribuição binomial:

velocidades implica que a função de repartição das velocidades, em qualquer direcção dada, de uma molécula de massa M de um gás, à temperatura absoluta T , é normal com média 0 e variância $M/(kT)$ com k constante (a constante de Boltzmann).

A distribuição normal é igualmente de importância teórica devido a poder usar-se de forma aproximada para as distribuições de muitos fenómenos aleatórios. Um exemplo é a distribuição binomial com parâmetros n e p . Se X é uma variável aleatória com distribuição binomial de parâmetros n e p , a probabilidade $P(a \leq X \leq b)$ é dada pela soma

$$\sum_{k=a}^b \binom{n}{k} p^k q^{n-k},$$

com $q = 1 - p$. Para grandes valores de n são necessários cálculos laboriosos para determinar esta soma. Na prática evitam-se estes cálculos recorrendo à fórmula aproximada

$$(14.25) \quad \sum_{k=a}^b \binom{n}{k} p^k q^{n-k} \sim \Phi\left(\frac{b - np + \frac{1}{2}}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np - \frac{1}{2}}{\sqrt{npq}}\right),$$

significando o símbolo \sim que os dois membros de (14.25) são assintoticamente iguais, ou, por outras palavras, o cociente do primeiro membro pelo segundo tende para o limite 1 quando $n \rightarrow \infty$. A relação limite expressa em (14.25) é um caso particular do

chamado *teorema limite central* da teoria das probabilidades. Este teorema (discutido mais em pormenor na Secção 14.30) justifica a importância teórica das distribuições normais.

A figura 14.11 ilustra a fórmula aproximada (14.25) e mostra que pode ser exacta para um valor de n suficientemente grande. As linhas a tracejado são as ordenadas da função massa probabilística p de uma distribuição binomial com os parâmetros $n = 10$ e $p = \frac{1}{5}$. Estas ordenadas são calculadas pela fórmula

$$p(t) = P(X = t) = \binom{10}{t} \left(\frac{1}{5}\right)^t \left(\frac{4}{5}\right)^{10-t} \quad \text{para } t = 0, 1, 2, \dots, 10.$$

As ordenadas para $t = 7, 8, 9$ e 10 não estão traçadas porque aí o valor de probabilidade é praticamente zero. Por exemplo, $p(10) = \left(\frac{1}{5}\right)^{10} = 2^{10}/10^{10} = 0,0000001024$. A curva regular é o gráfico da função densidade f de um distribuição normal (com média $m = np = 2$ e variância $\sigma^2 = npq = 1,6$). Para calcular a probabilidade $P(a \leq t \leq b)$ a partir da função massa probabilística p , adicionamos os valores da função $p(t)$ nos pontos em que está concentrada a massa do intervalo $a \leq t \leq b$. Cada valor de $p(t)$ pode ser interpretado como a área de um rectângulo de altura $p(t)$ localizado sobre um intervalo de comprimento unitário centrado em t . (Como exemplo, temos na figura 14.11 um rectângulo desses centrado em $t = 3$). A fórmula (14.25) resulta de se substituírem as áreas desses rectângulos pela área do conjunto de ordenadas de f referente ao intervalo $[a - \frac{1}{2}, b + \frac{1}{2}]$.

14.15. Indicações referentes a distribuições mais gerais

Nas secções precedentes discutimos exemplos de distribuições discretas e contínuas. Os valores da probabilidade de uma distribuição discreta calculam-se adicionando os valores da correspondente massa probabilística. Esses valores numa distribuição contínua calculam-se por integração da função densidade. Existem, todavia, distribuições que não são nem discretas nem contínuas. Entre estas estão as chamadas distribuições de tipo "misto", nas quais a distribuição da massa probabilística é parcialmente discreta e parcialmente contínua. (Na figura 14.3 mostra-se um exemplo).

Uma função de repartição F diz-se *mista* se pode exprimir-se por uma combinação linear da forma

$$(14.26) \quad F(t) = c_1 F_1(t) + c_2 F_2(t),$$

sendo F_1 discreta e F_2 contínua. As constantes c_1 e c_2 devem verificar as relações

$$0 < c_1 < 1, \quad 0 < c_2 < 1, \quad c_1 + c_2 = 1.$$

As propriedades das distribuições mistas podem estabelecer-se estudando as que são discretas ou contínuas e recorrendo à linearidade expressa na igualdade (14.26).

Um tipo de integral, conhecido como *integral de Riemann-Stieltjes*, torna possível o tratamento simultâneo dos casos discreto, contínuo e misto†. Embora este integral unifique o estudo teórico das funções de repartição, em cada problema específico a determinação das probabilidades reduz-se à soma e integração ordinárias. Nesta exposição introdutória não pretendemos fazer o estudo do integral de Riemann-Stieltjes. Por tal facto estudamos paralelamente a maior parte dos tópicos para o caso discreto e para o caso contínuo. Todavia, apresentamos os detalhes completos unicamente para um dos casos, deixando ao leitor o trabalho de os completar para outro caso.

Ainda mesmo o integral de Riemann-Stieltjes é inadequado para o tratamento de funções de repartição *mais gerais*. Mas um conceito mais forte, o chamado integral de Lebesgue-Stieltjes††, permite-nos tratar satisfatoriamente todos os casos. O estudo da teoria das probabilidades sob um ponto de vista superior não pode fazer-se sem um conhecimento perfeito do integral de Lebesgue-Stieltjes.

14.16. Exercícios

1. Seja X uma variável aleatória que mede o tempo de vida (em horas) de um certo tipo de válvula electrónica. Supor para X uma distribuição exponencial com parâmetro $\lambda = 0,001$. Determinar T de maneira que $P(X > T)$ seja:
(a) 0,90; (b) 0,99. Pode utilizar-se nos cálculos a fórmula aproximada $-\log(1-x) = x + x^2/2$.
2. Um material radioactivo desintegra-se segundo uma lei exponencial, com o período de vida média de 2 anos. Considerar o tempo de desintegração X (em anos) de um átomo e supor que X é uma variável aleatória com uma distribuição exponencial. Calcular a probabilidade de que um átomo se desintegre: (a) no intervalo $1 \leq X \leq 2$; (b) no intervalo $2 \leq X \leq 3$; (c) no intervalo $2 \leq X \leq 3$, sabendo que não se desintegrou no intervalo $0 \leq X \leq 2$; (d) no intervalo $2 \leq X \leq 3$, sabendo que não se desintegrou no intervalo $1 \leq X \leq 2$.
3. Verificou-se que a duração (em minutos) das chamadas interurbanas em determinada cidade, é um fenómeno aleatório com a função densidade de probabilidade

$$f(t) = \begin{cases} ce^{-t/3} & \text{para } t > 0, \\ 0 & \text{para } t \leq 0. \end{cases}$$

Determinar o valor de c e calcular a probabilidade de que uma chamada interurbana dure:
(a) menos que 3 minutos; (b) mais que 6 minutos; (c) entre 3 e 6 minutos; (d) mais do que 9 minutos.

4. Seja

$$f(t) = \begin{cases} \lambda e^{-\lambda(t-c)} & \text{se } t \geq c, \\ 0 & \text{se } t < c. \end{cases}$$

com c e λ conhecidos.

† No capítulo 9 do livro do autor, *Mathematical Analysis*, Addison-Wesley (1957) (tradução espanhola *Análisis Matemático*, Editorial Reverté) encontra-se um estudo da teoria do integral de Riemann-Stieltjes.

†† Ver qualquer livro sobre teoria da medida.

Verificar que $\int_{-\infty}^{\infty} f(t)dt = 1$ e determinar a função de repartição F tendo f como função densidade. Esta é a distribuição exponencial com dois parâmetros, um *parâmetro de desintegração* c e um *parâmetro de posição* λ .

- Estabelecer e demonstrar uma extensão do teorema 14.9 para distribuições exponenciais com dois parâmetros λ e c .
- Uma variável aleatória X tem uma distribuição exponencial com dois parâmetros λ e c . Seja $Y = aX + b$, com $a > 0$. Provar que Y tem também uma distribuição exponencial com dois parâmetros λ' e c' e determinar estes parâmetros em função de a , b , c e λ .
- No Exercício 16 da Secção 11.28 provou-se que $\int_0^{\infty} e^{-x^2} dx = \sqrt{\pi}/2$. Utilizar este resultado para demonstrar que, para $\sigma > 0$, tem-se

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left(\frac{u-m}{\sigma}\right)^2\right\} du = 1.$$

- Uma variável aleatória X tem uma distribuição normal standard Φ . Provar que: (a) $\Phi(-x) = 1 - \Phi(x)$; (b) $P(|X| < k) = 2\Phi(k) - 1$; (c) $P(|X| > k) = 2(1 - \Phi(k))$.
- Uma variável aleatória X tem uma distribuição normal standard Φ . Utilizar a tábua 14.1 para calcular cada uma das seguintes probabilidades: (a) $P(X > 0)$; (b) $P(1 < X < 2)$; (c) $P(X < 3)$; (d) $P(|X| > 2)$.
- Uma variável aleatória X tem uma distribuição normal standard Φ . Utilizar a tábua 14.1 para determinar c tal que: (a) $P(|X| > c) = \frac{1}{2}$; (b) $P(|X| > c) = 0,98$.
- Supor que X tem uma função de repartição normal F com média m e variância σ^2 , e seja Φ a distribuição normal standard correspondente.
(a) Provar que

$$F(t) = \Phi\left(\frac{t-m}{\sigma}\right).$$

- (b) Determinar um valor de c tal que $P(|X - m| > c) = \frac{1}{2}$.
- (c) Determinar um valor de c tal que $P(|X - m| > c) = 0,98$.
- Uma variável aleatória X tem uma distribuição normal com média $m = 1$ e variância $\sigma^2 = 4$. Calcular cada uma das seguintes probabilidades: (a) $P(-3 \leq X \leq 3)$; (b) $P(-5 \leq X \leq 3)$.
- Um arquiteto desenha um portal de um edifício a ser utilizado por pessoas cuja altura está distribuída normalmente, com média $m = 1,8$ metros e variância σ^2 , com $\sigma = 10$ centímetros. Qual a altura mínima possível para o portal de modo a que não mais do que 1% das pessoas toque com a cabeça na parte superior do mesmo?
- Se X tem uma distribuição normal standard, provar que a variável aleatória $Y = aX + b$ é também normal se $a \neq 0$. Determinar a média e a variância de Y .
- Supor uma variável aleatória X com uma distribuição normal standard, e seja $Y = X^2$.
(a) Provar que $F_Y(t) = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{t}} e^{-u^2/2} du$ se $t \geq 0$.
- (b) Determinar $F_Y(t)$ quando $t < 0$ e descrever a função densidade f_Y .

14.17. Distribuições de funções de variáveis aleatórias

Se φ é uma função real cujo domínio inclui o contradomínio da variável aleatória X , podemos construir uma nova variável Y pela equação

$$Y = \varphi(X),$$

o que significa que $Y(\omega) = \varphi[X(\omega)]$ para cada ω do conjunto fundamental. Se conhecermos a função de repartição F_X de X , como poderemos determinar a função de repartição F_Y de Y ? Começamos por apresentar um caso particular importante. Suponhamos que φ é contínua e estritamente crescente em todo o eixo real. Neste caso φ admite uma função inversa ψ , contínua e estritamente crescente, tal que, para cada x e y ,

$$y = \varphi(x) \quad \text{se, e só se,} \quad x = \psi(y).$$

Pela definição de F_Y temos

$$F_Y(t) = P(Y \leq t) = P[\varphi(X) \leq t].$$

Uma vez que φ é estritamente crescente e contínua, os acontecimentos " $\varphi(X) \leq t$ " e " $X \leq \psi(t)$ " são idênticos. Deste modo $P[\varphi(X) \leq t] = P[X \leq \psi(t)] = F_X[\psi(t)]$. Logo as distribuições F_Y e F_X estão relacionadas pela equação

$$(14.27) \quad F_Y(t) = F_X[\psi(t)].$$

Quando a função de repartição F_X e a função ψ admitem derivadas podemos derivar ambos os membros de (14.27), aplicando no segundo membro a regra da derivada da função composta, para obtermos

$$F'_Y(t) = F'_X[\psi(t)] \cdot \psi'(t).$$

Esta igualdade proporciona-nos a seguinte relação entre as densidades:

$$f_Y(t) = f_X[\psi(t)] \cdot \psi'(t).$$

EXEMPLO 1. $Y = aX + b$, $a > 0$. Neste caso temos

$$\varphi(x) = ax + b, \quad \psi(y) = \frac{y-b}{a}, \quad \psi'(y) = \frac{1}{a}.$$

Visto φ ser contínua e estritamente crescente podemos escrever

$$F_Y(t) = F_X\left(\frac{t-b}{a}\right) \quad \text{e} \quad f_Y(t) = \frac{1}{a} f_X\left(\frac{t-b}{a}\right).$$

EXEMPLO 2. $Y = X^2$. Neste caso $\varphi(x) = x^2$ e o raciocínio atrás efectuado não é directamente aplicável porque φ não é aqui estritamente crescente. Todavia, podemos utilizar o mesmo tipo de raciocínio para determinar F_Y e f_Y . Pela definição de F_Y temos

$$F_Y(t) = P(X^2 \leq t).$$

Se $t < 0$ o acontecimento " $X^2 \leq t$ " é vazio e por isso $P(X^2 \leq t) = 0$. Deste modo $F_Y(t) = 0$ para $t < 0$. Se $t > 0$ temos

$$P(X^2 \leq t) = P(-\sqrt{t} \leq X \leq \sqrt{t}) = F_X(\sqrt{t}) - F_X(-\sqrt{t}) + P(X = -\sqrt{t}).$$

Para uma distribuição contínua F_X temos $P(X = -\sqrt{t}) = 0$ e resulta a seguinte relação entre F_Y e F_X :

$$F_Y(t) = \begin{cases} 0 & \text{se } t < 0, \\ F_X(\sqrt{t}) - F_X(-\sqrt{t}) & \text{se } t > 0. \end{cases}$$

Para todo o $t < 0$ e para aqueles valores de $t > 0$ tais que F_X é derivável em \sqrt{t} e em $-\sqrt{t}$ temos as seguintes relações entre as densidades:

$$f_Y(t) = \begin{cases} 0 & \text{se } t < 0, \\ \frac{f_X(\sqrt{t}) + f_X(-\sqrt{t})}{2\sqrt{t}} & \text{se } t > 0. \end{cases}$$

Na Secção 14.23 serão analisados outros problemas deste tipo com o auxílio de variáveis aleatórias bidimensionais.

14.18. Exercícios

- Supor que X tem uma distribuição uniforme no intervalo $[0, 1]$. Determinar a função de repartição F_Y e a densidade de probabilidade f_Y da variável aleatória Y se:
 - $Y = 3X + 1$,
 - $Y = -3X + 1$,
 - $Y = X^2$,
 - $Y = \log |X|$,
 - $Y = \log X^2$,
 - $Y = e^X$.
- Seja X uma variável aleatória admitindo uma função de repartição contínua F_X . Se φ é contínua e estritamente crescente em todo o eixo real e se $\varphi(x) \rightarrow a$ quando $x \rightarrow -\infty$ e $\varphi(x) \rightarrow b$ quando $x \rightarrow +\infty$, determinar a função de repartição F_Y da variável aleatória $Y = \varphi(X)$. Determinar também a densidade f_Y , admitindo que F_X e φ são deriváveis.
- Supor que X tem uma distribuição normal standard. Determinar uma função densidade de probabilidade da variável aleatória Y quando
 - $Y = X^2$,
 - $Y = |X|^{1/2}$,
 - $Y = e^X$,
 - $Y = \arctan X$.

14.19. Distribuição de variáveis aleatórias bidimensionais

O conceito de distribuição pode generalizar-se para variáveis aleatórias n -dimensionais de uma maneira directa. O estudo do caso $n = 2$ indicar-nos-á o modo como essa generalização é feita.

Se X e Y são duas variáveis aleatórias unidimensionais definidas num mesmo conjunto fundamental S , (X, Y) representará a variável aleatória bidimensional cujo va-

ior num ponto genérico ω de S é dado pelo par de números reais $(X(\omega), Y(\omega))$. A notação

$$X \leq a, Y \leq b$$

é uma forma abreviada para designar o conjunto de todos os elementos ω de S tais que $X(\omega) \leq a$ e $Y(\omega) \leq b$; a probabilidade deste acontecimento representa-se por

$$P(X \leq a, Y \leq b).$$

Analogamente se interpretam as notações $a < X \leq b, c < Y \leq d$, e $P(a < X \leq b, c < Y \leq d)$.

O conjunto de pontos (x, y) tais que $x \leq a$ e $y \leq b$ é o produto cartesiano $A \times B$ dos dois intervalos infinitos $A = \{x | x \leq a\}$ e $B = \{y | y \leq b\}$. O conjunto $A \times B$ representa-se geometricamente pela região rectangular infinita representada na fig. 14.12. O número $P(X \leq a, Y \leq b)$ representa a probabilidade de que um ponto $(X(\omega), Y(\omega))$ esteja nesta região. Estas são as probabilidades bidimensionais análogas às probabilidades unidimensionais $P(X \leq a)$ e utilizam-se para definir as distribuições de probabilidades bidimensionais.

DEFINIÇÃO. A função de repartição da variável aleatória bidimensional (X, Y) é a função real F definida para todo o par de números reais a e b pela igualdade

$$F(a, b) = P(X \leq a, Y \leq b).$$

É também designada por distribuição conjunta das duas variáveis aleatórias unidimensionais X e Y .

Para calcular a probabilidade de que (X, Y) esteja num rectângulo recorreremos ao teorema seguinte que é uma generalização do teorema 14.2(b)

TEOREMA 14.10. Se F é a função de repartição de uma variável aleatória bidimensional (X, Y) , então se $a < b$ e $c < d$ tem-se

$$(14.28) \quad P(a < X \leq b, c < Y \leq d) = F(b, d) - F(a, d) - F(b, c) + F(a, c).$$

Demonstração. Os dois acontecimentos " $X \leq a, c < Y \leq d$ " e " $X \leq a, Y \leq c$ " são disjuntos e a sua união é " $X \leq a, Y \leq d$ ". Adicionando as probabilidades obtemos $P(X \leq a, c < Y \leq d) + P(X \leq a, Y \leq c) = P(X \leq a, Y \leq d)$; logo

$$P(X \leq a, c < Y \leq d) = F(a, d) - F(a, c).$$

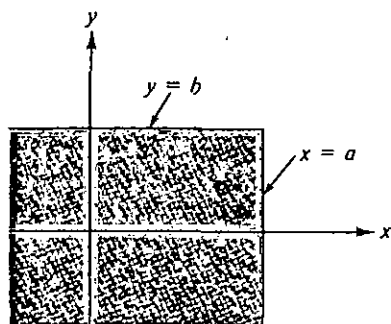


FIG. 14.12. Uma região rectangular infinita $A \times B$, onde $A = \{x \mid x \leq a\}$ e $B = \{y \mid y \leq b\}$.

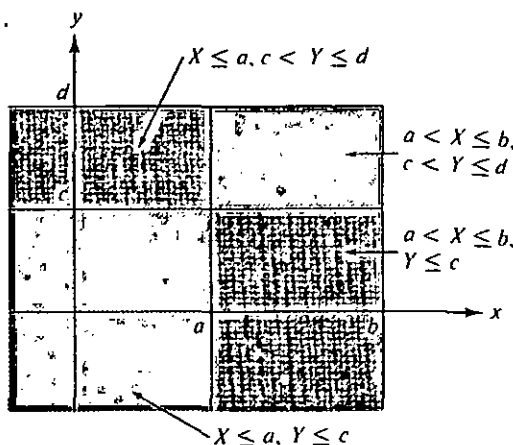


FIG. 14.13. O acontecimento " $X \leq b, Y \leq d$ " expresso como a união de quatro acontecimentos disjuntos.

Analogamente, temos

$$P(a < X \leq b, Y \leq c) = F(b, c) - F(a, c).$$

Então os quatro acontecimentos

$$"X \leq a, Y \leq c,"$$

$$"X \leq a, c < Y \leq d,"$$

$$"a < X \leq b, Y \leq c,"$$

$$"a < X \leq b, c < Y \leq d"$$

são disjuntos, e a sua união é " $X \leq b, Y \leq d$ ". (Ver figura 14.13.) Somando as correspondentes probabilidades e considerando as igualdades precedentes obtemos

$$F(a, c) + [F(a, d) - F(a, c)] + [F(b, c) - F(a, c)] + P(a < X \leq b, c < Y \leq d) = F(b, d)$$

o que é equivalente a (14.28).

A fórmula (14.28) dá-nos a probabilidade de que a variável aleatória (X, Y) tenha um valor no rectângulo $(a, b) \times (c, d)$. Existem, naturalmente, fórmulas análogas para os rectângulos, $[a, b] \times [c, d]$, $(a, b) \times (c, d)$, $[a, b] \times (c, d)$, etc.

Nota: A analogia com a distribuição de massa pode generalizar-se ao caso bi-dimensional. Aqui a massa total 1 distribui-se numa determinada região plana. A probabilidade $P(a < X \leq b, c < Y \leq d)$ representa a quantidade de massa probabilística situada no rectângulo $(a, b) \times (c, d)$. O número $F(a, b)$ mede a massa

pertencente à região rectangular infinita $X \leq a$, $Y \leq b$. Como no caso unidimensional, os dois tipos mais importantes de distribuições são designados por *discretas* e *contínuas*. No caso discreto toda a massa probabilística está concentrada num número finito ou numa infinidade numerável de pontos. No caso contrário a massa é distribuída por toda uma região plana, distribuição essa feita de maneira uniforme ou variável.

14.20. Distribuições discretas bidimensionais

Dada a variável aleatória (X, Y) , podemos definir uma nova função p , chamada a função massa probabilística de (X, Y) , tal que

$$p(x, y) = P(X = x, Y = y)$$

para todo o par de números reais (x, y) . Seja T o conjunto dos pontos (x, y) para os quais $p(x, y) > 0$. Pode provar-se que T é finito ou infinito numerável. Se a soma dos valores $p(x, y)$ para todo (x, y) em T é igual a 1, isto é, se

$$(14.29) \quad \sum_{(x, y) \in T} p(x, y) = 1,$$

a variável aleatória (X, Y) diz-se *discreta*. Os pontos (x, y) de T chamam-se *pontos de massa* (probabilística) de (X, Y) .

Suponhamos que x_1, x_2, x_3, \dots , e y_1, y_2, y_3, \dots , são valores possíveis de X e Y , respectivamente, e seja

$$p_{ij} = P(X = x_i, Y = y_j).$$

Se cada p_{ij} for positivo e a soma de todos os p_{ij} for 1, então a probabilidade de um acontecimento " $(X, Y) \in E$ " é a soma de todos os p_{ij} considerando todos os x_i e todos os y_j para os quais $(x_i, y_j) \in E$. Indicamos isto escrevendo

$$P[(X, Y) \in E] = \sum_{\substack{x_i, y_j \\ (x_i, y_j) \in E}} p_{ij}.$$

Em particular, visto ser $P(X \leq x, Y \leq y) = F(x, y)$, a distribuição conjunta F (que também se chama discreta) é dada pela dupla soma

$$(14.30) \quad F(x, y) = \sum_{x_i \leq x} \sum_{y_j \leq y} p_{ij}.$$

Os valores p_{ij} podem também usar-se para reconstruir as funções massa probabilística p_X e p_Y das variáveis aleatórias unidimensionais X e Y . Com efeito, se E_{ij} representa o acontecimento " $X = x_i, Y = y_j$ ", os acontecimentos $E_{i1}, E_{i2}, E_{i3}, \dots$ são disjuntos e a sua união é o acontecimento " $X = x_i$ ". Logo, pela aditividade numerável, obtemos

$$(14.31) \quad P(X = x_i) = \sum_{j=1}^{\infty} P(E_{ij}) = \sum_{j=1}^{\infty} p_{ij}.$$

Analogamente, temos

$$(14.32) \quad P(Y = y_j) = \sum_{i=1}^{\infty} P(E_{ij}) = \sum_{i=1}^{\infty} p_{ij}.$$

Deste modo as correspondentes distribuições unidimensionais F_X e F_Y podem calcular-se por intermédio das fórmulas

$$F_X(t) = \sum_{x_i \leq t} P(X = x_i) = \sum_{x_i \leq t} \sum_{j=1}^{\infty} p_{ij}$$

e

$$F_Y(t) = \sum_{y_j \leq t} P(Y = y_j) = \sum_{y_j \leq t} \sum_{i=1}^{\infty} p_{ij}.$$

Para conjuntos fundamentais finitos, naturalmente, as séries devem ser substituídas por somas finitas.

14.21. Distribuições bidimensionais contínuas. Funções densidade

Como era natural supôr, as *distribuições contínuas* são as que são contínuas em todo o plano. Para a maior parte das funções de repartição contínuas F que aparecem na prática é possível definir uma função não negativa f (chamada a *densidade de probabilidade* de F) tal que as probabilidades de muitos acontecimentos de interesse podem calcular-se por dupla integração da função densidade. Por outras palavras, a probabilidade de um acontecimento " $(X, Y) \in Q$ " é dada pelo integral

$$(14.33) \quad P[(X, Y) \in Q] = \iint_Q f.$$

Quando uma tal função f existe chama-se também densidade de probabilidade de variável aleatória (X, Y) , ou densidade conjunta de X e Y . Não tentaremos descrever a classe de regiões Q para as quais (14.33) é válida excepto para mencionarmos que esta classe será suficientemente ampla para incluir todas as regiões que se apresentam nas aplicações usuais das probabilidades. Por exemplo, se existe uma densidade conjunta temos sempre

$$(14.34) \quad P(a < X \leq b; c < Y \leq d) = \iint_R f(x, y) dx dy,$$

em que $R = [a, b] \times [c, d]$. A função integranda f é habitualmente de natureza tal que o integral duplo pode calcular-se por intermédio de duas integrações unidimensionais sucessivas, caso em que (14.34) se escreve

$$P(a < X \leq b, c < Y \leq d) = \int_c^d \left[\int_a^b f(x, y) dx \right] dy = \int_a^b \left[\int_c^d f(x, y) dy \right] dx.$$

Em todos os exemplos que considerarmos, esta fórmula é também válida nos casos limites em que a e c se substituem por $-\infty$ e b e d por $+\infty$. Assim temos

$$(14.35) \quad F(b, d) = \int_{-\infty}^d \left[\int_{-\infty}^b f(x, y) dx \right] dy = \int_{-\infty}^b \left[\int_{-\infty}^d f(x, y) dy \right] dx$$

quaisquer que sejam b e d , e

$$(14.36) \quad \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} f(x, y) dx \right] dy = \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} f(x, y) dy \right] dx = 1.$$

As igualdades (14.35) e (14.36) são as correspondentes, no caso contínuo, a (14.30) e (14.29), respectivamente.

Se a função densidade existe não é única, visto que o integrando de (14.33) pode mudar-se num número finito de pontos sem afectar o valor do integral. Todavia existe quando muito uma função densidade contínua. Com efeito, nos pontos de continuidade de f temos as fórmulas

$$f(x, y) = D_{1,2}F(x, y) = D_{2,1}F(x, y),$$

obtidas por derivação dos integrais em (14.35).

Como no caso discreto, a densidade conjunta f pode usar-se na determinação das densidades unidimensionais f_X e f_Y . As fórmulas análogas a (14.31) e (14.32) são

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy \quad \text{e} \quad f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx.$$

As correspondentes funções de repartição $F_X(t)$ e $F_Y(t)$ obtêm-se, naturalmente, por integração das respectivas densidades f_X e f_Y de $-\infty$ a t .

As variáveis aleatórias X e Y dizem-se *independentes* se a função de repartição $F(x, y)$ goza da propriedade de

$$F(x, y) = F_X(x)F_Y(y)$$

para todo o par (x, y) . Na próxima Secção de Exercícios discutiremos algumas consequências da independência de X e Y .

EXEMPLO. Considerar a função f com o valor constante 1 sobre o quadrado $R = [0, 1] \times [0, 1]$, e o valor 0 em todo o restante plano. Uma variável aleatória (X, Y) tendo esta função densidade diz-se estar *uniformemente distribuída* sobre R . A correspondente função de repartição F é definida por:

$$F(x, y) = \begin{cases} xy & \text{se } (x, y) \in R, \\ x & \text{se } 0 < x < 1 \quad \text{e} \quad y > 1, \\ y & \text{se } 0 < y < 1 \quad \text{e} \quad x > 1, \\ 1 & \text{se } x \geq 1 \quad \text{e} \quad y \geq 1, \\ 0 & \text{para os restantes pontos} \end{cases}$$

O gráfico de F sobre R é uma parte da superfície $z = xy$ (sela de cavalo). Em todos os pontos (x, y) não situados sobre fronteira de R as derivadas parciais mistas $D_{1,2}F(x, y)$ e $D_{2,1}F(x, y)$ existem e são iguais a $f(x, y)$. Esta distribuição é o produto de duas distribuições uniformes unidimensionais F_X e F_Y . Logo X e Y são independentes.

14.22. Exercícios

- Sejam X e Y variáveis aleatórias unidimensionais com funções de repartição F_X e F_Y e F a função de repartição da distribuição conjunta de X e Y .
(a) Provar que X e Y são independentes se e só se

$$P(a < X \leq b, c < Y \leq d) = P(a < X \leq b)P(c < Y \leq d)$$

quaisquer que sejam a, b, c, d , com $a < b$ e $c < d$.

- Considerar o caso discreto. Supôr x_1, x_2, \dots e y_1, y_2, \dots como os pontos de massa de X e Y , respectivamente. Seja $a_i = P(X = x_i)$ e $b_j = P(Y = y_j)$. Se $p_{ij} = P(X = x_i, Y = y_j)$ mostrar que X e Y são independentes se $p_{ij} = a_i b_j$ quaisquer que sejam i e j .
- Sejam X e Y variáveis aleatórias com distribuições contínuas com as correspondentes densidades f_X e f_Y e represente f a densidade da distribuição conjunta. Supôr contínuas as três densidades. Provar que a condição de independência é equivalente à afirmação de que $f(x, y) = f_X(x)f_Y(y)$ para todo o par (x, y) . [Sugestão: Expressar f como derivada da função de repartição F .]
- Considerar o Exercício 1. Supôr que $P(X = x_1, Y = y_1) = P(X = x_2, Y = y_2) = p/2$ e que $P(X = x_1, Y = y_2) = P(X = x_2, Y = y_1) = q/2$, com p e q números não negativos de soma 1.
 - Determinar as probabilidades $P(X = x_i)$ e $P(Y = y_j)$ para $i = 1, 2$ e $j = 1, 2$.
 - Para que valor (ou valores) de p serão independentes X e Y ?
- Se $a < b$ e $c < d$, define-se f do modo seguinte

$$f(x, y) = \begin{cases} \frac{1}{(b-a)(d-c)} & \text{se } (x, y) \in [a, b] \times [c, d], \\ 0 & \text{para os restantes pontos.} \end{cases}$$

- Verificar que esta é a densidade de uma distribuição contínua F e determinar F .
- Determinar as distribuições unidimensionais F_X e F_Y .
- Determinar se X e Y são ou não independentes.
- Se $P(Y \leq b) \neq 0$, a probabilidade condicionada de que $X \leq a$, sabido que $Y \leq b$, representa-se por $P(X \leq a \mid Y \leq b)$, e define-se pela fórmula

$$P(X \leq a | Y \leq b) = \frac{P(X \leq a, Y \leq b)}{P(Y \leq b)}.$$

Se $P(Y \leq b) = 0$, definimos $P(X \leq a | Y \leq b) = P(X \leq a)$. Analogamente, se $P(X \leq a) \neq 0$, definimos $P(Y \leq b | X \leq a) = P(X \leq a, Y \leq b) / P(X \leq a)$. Se $P(X \leq a) = 0$, definimos $P(Y \leq b | X \leq a) = P(Y \leq b)$.

(a) Considerar o Exercício 1 e provar a independência de X e Y em função das probabilidades condicionadas.

(b) Considerar o caso discreto. Supor x_1, x_2, \dots , e y_1, y_2, \dots pontos de massa de X e Y , respectivamente. Provar que

$$P(X = x_i) = \sum_{j=1}^{\infty} P(Y = y_j) P(X = x_i | Y = y_j)$$

e

$$P(Y = y_j) = \sum_{i=1}^{\infty} P(X = x_i) P(Y = y_j | X = x_i).$$

5. Uma casa de jogo oferece aos seus clientes o seguinte jogo: Lança-se uma moeda ao ar. Se o resultado do primeiro lançamento for face “cruz”, o jogador perde e o jogo *acaba*. Se o resultado do primeiro lançamento for face “cara”, efectua-se novo lançamento. Se sai “cara” da segunda vez o jogador ganha 2\$00, mas se sai “cruz” o jogador ganha unicamente 1\$00. Seja X a variável aleatória que toma os valores 1 ou 0, conforme no primeiro lançamento sai “cara” ou “cruz”. Seja Y a variável aleatória que dá conta do número de escudos ganhos pelo jogador. Recorrer ao Exercício 4 (ou a qualquer outro método) para calcular $P(Y = 0)$, $P(Y = 1)$, e $P(Y = 2)$.

6. Considerar o Exercício 4. Deduzir as chamadas fórmulas de Bayer

$$P(X = x_k | Y = y_j) = \frac{P(X = x_k) P(Y = y_j | X = x_k)}{\sum_{i=1}^{\infty} P(X = x_i) P(Y = y_j | X = x_i)},$$

$$P(Y = y_k | X = x_i) = \frac{P(Y = y_k) P(X = x_i | Y = y_k)}{\sum_{j=1}^{\infty} P(Y = y_j) P(X = x_i | Y = y_j)}.$$

7. Dadas duas urnas A e B , a urna A contém uma nota de 50\$00 e duas de 100\$00 e a urna B contém três notas de 50\$00 e uma de 100\$00. Extrai-se uma nota da urna A e introduz-se em B . Seja Y a variável aleatória que dá conta do valor da nota mudada. Extrai-se depois uma nota da urna B e define-se a variável aleatória X para representar o seu valor. Calcular as probabilidades condicionadas

$$P(Y = 5 | X = 10) \quad \text{e} \quad P(Y = 10 | X = 10).$$

[Sugestão: Aplicar as fórmulas de Bayes do Exercício 6].

8. São dadas três caixas idênticas, cada uma com duas gavetas. A caixa número 1 tem uma moeda de ouro numa gaveta e uma de prata na outra. A caixa número 2 tem uma moeda de ouro em cada gaveta e a caixa número 3 tem uma moeda de prata em cada gaveta. Abre-se ao acaso uma gaveta e verifica-se que contém 1 moeda de ouro. Calcular a probabilidade de que na outra gaveta da mesma caixa exista uma moeda de prata. [Sugestão: Aplicar as fórmulas de Bayes do Exercício 6].

9. Seja Q uma região plana e $a(Q)$ a sua área (positiva). Uma variável aleatória bidimensional (X, Y) diz-se ter uma *distribuição uniforme* sobre Q se a correspondente função densidade se define do modo seguinte

$$f(x, y) = \begin{cases} 1/a(Q) & \text{se } (x, y) \in Q, \\ 0 & \text{se } (x, y) \notin Q. \end{cases}$$

- (a) Se E é uma subregião de Q com área $a(E)$, provar que $a(E)/a(Q)$ é a probabilidade do acontecimento $(X, Y) \in E$.
 (b) Caem ao acaso gotas de água sobre o quadrado Q de vértices $(1, 0)$, $(0, 1)$, $(-1, 0)$, $(0, -1)$. Um resultado é o ponto (x, y) de Q atingido por uma determinada gota. Seja $X(x, y) = x$ e $Y(x, y) = y$ e admita-se que (X, Y) tem uma distribuição uniforme sobre Q . Determinar a função densidade conjunta f e as densidades unidimensionais f_X e f_Y . Serão as variáveis aleatórias X e Y independentes?
 10. Uma variável aleatória bidimensional (X, Y) tem a função de repartição F . Seja $U = X - a$ e $V = Y - b$, com a e b constantes. Se G representa a distribuição conjunta de (U, V) provar que

$$G(u, v) = F(u + a, v + b).$$

Deduzir uma relação semelhante ligando as funções densidade f de (X, Y) e g de (U, V) quando f é contínua.

14.23. Distribuição de funções de duas variáveis aleatórias

Fixamos agora a nossa atenção no problema seguinte: Se X e Y são variáveis aleatórias unidimensionais com distribuições conhecidas, como determinar a distribuição de novas variáveis aleatórias tais como $X + Y$, XY , ou $X^2 + Y^2$? Vamos aqui expôr um método que nos permite responder a perguntas deste tipo. Definimos duas novas variáveis U e V por relações da forma

$$U = M(X, Y), \quad V = N(X, Y),$$

onde $M(X, Y)$ ou $N(X, Y)$ é a combinação particular em que estamos interessados. A partir do conhecimento da distribuição conjunta f da variável aleatória bidimensional (X, Y) calculamos a distribuição conjunta g de (U, V) . Uma vez que se conhece g , as distribuições de U e V são facilmente determináveis.

Para descrever o método em pormenor, consideremos uma aplicação biunívoca do plano xy sobre o plano uv definida pelo par de equações

$$u = M(x, y), \quad v = N(x, y).$$

Seja a aplicação inversa definida por

$$x = Q(u, v), \quad y = R(u, v),$$

e admitamos que Q e R possuem derivadas parciais contínuas. Se T representa uma região do plano xy , seja T' a sua imagem no plano uv , como se sugere na figura 14.14.

Representemos por X e Y duas variáveis aleatórias contínuas unidimensionais, tendo uma distribuição conjunta e suponhamos que (X, Y) tem uma função densidade de probabilidade f . Definamos as novas variáveis aleatórias U e V escrevendo $U = M(X, Y)$, $V = N(X, Y)$. Para determinar uma função densidade de probabilidade g da variável aleatória (U, V) procedemos do modo seguinte:

As variáveis aleatórias X e Y estão associadas a um conjunto fundamental S . Para cada ω de S , temos $U(\omega) = M[X(\omega), Y(\omega)]$ e $V(\omega) = N[X(\omega), Y(\omega)]$. Uma vez que a aplicação é biunívoca, os dois conjuntos

$$\{\omega \mid (U(\omega), V(\omega)) \in T'\} \quad \text{e} \quad \{\omega \mid (X(\omega), Y(\omega)) \in T\}$$

são iguais. Deste modo temos

$$(14.37) \quad P[(U, V) \in T'] = P[(X, Y) \in T].$$

Uma vez que f é a função densidade de (X, Y) podemos escrever

$$(14.38) \quad P[(X, Y) \in T] = \iint_T f(x, y) dx dy.$$

Considerando (14.37) e a fórmula de transformação do integral duplo podemos escrever (14.38) do modo seguinte:

$$P[(U, V) \in T'] = \iint_{T'} f[Q(u, v), R(u, v)] \left| \frac{\partial(Q, R)}{\partial(u, v)} \right| du dv.$$

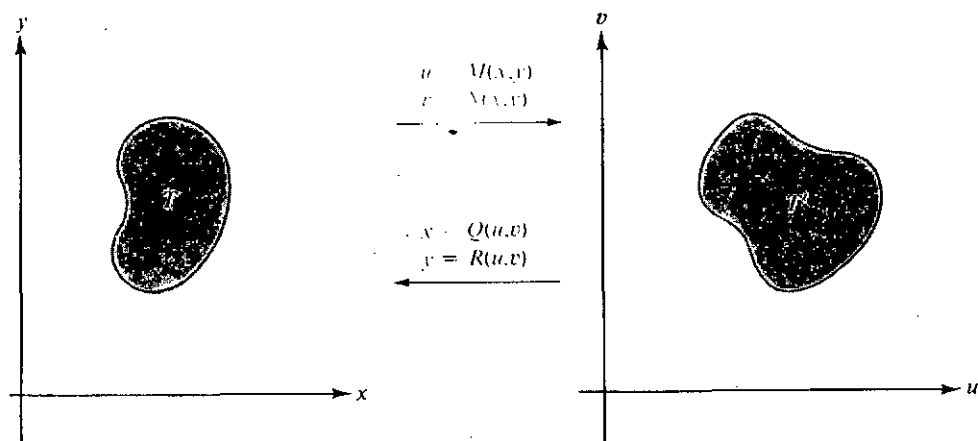


FIG. 14.14. Aplicação biunívoca da região T do plano xy sobre a região T' do plano uv .

Porque isto é válido para cada região T do plano uv , uma função densidade g de (U, V) é dada pelo integrando do segundo membro; isto é, temos

$$(14.39) \quad g(u, v) = f[Q(u, v), R(u, v)] \left| \frac{\partial(Q, R)}{\partial(u, v)} \right|.$$

As densidades f_U e f_V podem agora obter-se pelas fórmulas

$$f_U(u) = \int_{-\infty}^{\infty} g(u, v) dv, \quad f_V(v) = \int_{-\infty}^{\infty} g(u, v) du.$$

EXEMPLO 1. *Soma e diferença de duas variáveis aleatórias.* Dadas duas variáveis aleatórias unidimensionais X e Y com a densidade conjunta f , determinar as funções densidade para as variáveis aleatórias $U = X + Y$, $V = X - Y$.

Resolução. Consideramos a aplicação definida por $u = x + y$, $v = x - y$. Esta transformação é não singular e consequentemente existe a transformação inversa definida por

$$x = \frac{u+v}{2} = Q(u, v), \quad y = \frac{u-v}{2} = R(u, v).$$

O determinante jacobiano é

$$\frac{\partial(Q, R)}{\partial(u, v)} = \begin{vmatrix} \frac{\partial Q}{\partial u} & \frac{\partial Q}{\partial v} \\ \frac{\partial R}{\partial u} & \frac{\partial R}{\partial v} \end{vmatrix} = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{2}.$$

Considerando (14.39) vemos que a densidade conjunta g de (U, V) é dada por

$$g(u, v) = f\left(\frac{u+v}{2}, \frac{u-v}{2}\right) \left| \frac{\partial(Q, R)}{\partial(u, v)} \right| = \frac{1}{2} f\left(\frac{u+v}{2}, \frac{u-v}{2}\right).$$

Para obter uma densidade $f_U = f_{X+Y}$ integramos relativamente a v e encontramos

$$f_{X+Y}(u) = \frac{1}{2} \int_{-\infty}^{\infty} f\left(\frac{u+v}{2}, \frac{u-v}{2}\right) dv.$$

A mudança de variável $x = \frac{1}{2}(u+v)$, $dx = \frac{1}{2}dv$, transforma esta na fórmula seguinte

$$f_{X+Y}(u) = \int_{-\infty}^{\infty} f(x, u-x) dx.$$

Do mesmo modo, encontramos

$$f_{X-Y}(v) = \frac{1}{2} \int_{-\infty}^{\infty} f\left(\frac{u+v}{2}, \frac{u-v}{2}\right) du = \int_{-\infty}^{\infty} f(x, x-v) dx.$$

Um caso particular importante ocorre quando X e Y são independentes. Neste caso a densidade conjunta exprime-se segundo o produto

$$f(x, y) = f_X(x)f_Y(y),$$

e os integrais para f_{X+Y} e f_{X-Y} escrevem-se

$$f_{X+Y}(u) = \int_{-\infty}^{\infty} f_X(x)f_Y(u-x) dx, \quad f_{X-Y}(v) = \int_{-\infty}^{\infty} f_X(x)f_Y(x-v) dx.$$

EXEMPLO 2. *Soma de duas distribuições exponenciais.* Suponhamos agora que cada uma das variáveis aleatórias X e Y tem uma distribuição exponencial, seja $f_X(t) = f_Y(t) = 0$ para $t < 0$, e

$$f_X(t) = \lambda e^{-\lambda t}, \quad f_Y(t) = \mu e^{-\mu t} \quad \text{para } t \geq 0.$$

Determinar a densidade de $X + Y$ quando X e Y são independentes.

Resolução. Se $u < 0$ o integral para $f_{X+Y}(u)$ é 0 porque o factor $f_X(x) = 0$ para $X < 0$, e o factor $f_Y(u-x) = 0$ para $x \geq 0$. Se $u \geq 0$ o integral que determina $f_{X+Y}(u)$ vem

$$f_{X+Y}(u) = \int_0^u \lambda e^{-\lambda x} \mu e^{-\mu(u-x)} dx = \lambda \mu e^{-\mu u} \int_0^u e^{(\mu-\lambda)x} dx.$$

Para calcular o último integral consideremos dois casos, $\mu = \lambda$ e $\mu \neq \lambda$.

Se $\mu = \lambda$ o integral tem o valor u e obtemos

$$f_{X+Y}(u) = \lambda^2 u e^{-\lambda u} \quad \text{para } u \geq 0.$$

Se $\mu \neq \lambda$ obtemos

$$f_{X+Y}(u) = \lambda \mu e^{-\mu u} \frac{e^{(\mu-\lambda)u} - 1}{\mu - \lambda} = \lambda \mu \frac{e^{-\lambda u} - e^{-\mu u}}{\mu - \lambda} \quad \text{para } u \geq 0.$$

EXEMPLO 3. *Máximo e mínimo de duas variáveis aleatórias independentes.* Sejam X e Y duas variáveis aleatórias unidimensionais com densidades f_X e f_Y e funções de repartição F_X e F_Y . Sejam U e V as variáveis aleatórias

$$U = \max\{X, Y\}, \quad V = \min\{X, Y\}.$$

Quer dizer, para cada ω do conjunto fundamental, $U(\omega)$ é o máximo e $V(\omega)$ é o mínimo dos dois números $X(\omega)$, $Y(\omega)$. A aplicação $u = \max\{x, y\}$, $v = \min\{x, y\}$ não é biunívoca, pelo que o método usado para deduzir a equação (14.39) não é aplicável. Todavia, neste caso podemos obter as funções de repartição de U e V directamente.

Em primeiro lugar observemos que $U \leq t$ se, e só se, $X \leq t$ e $Y \leq t$. Deste $P(U \leq t) = P(X \leq t, Y \leq t)$. Devido à independência esta igualdade pode escrever-se $P(U \leq t) = P(X \leq t) \cdot P(Y \leq t) = F_X(t)F_Y(t)$. Assim, temos

$$F_U(t) = F_X(t)F_Y(t).$$

Em cada ponto em que f_X e f_Y são contínuas podemos derivar esta igualdade para obtermos

$$f_U(t) = f_X(t)f_Y(t) + F_X(t)f_Y(t).$$

Analogamente, temos $V > t$ se e só se $X > t$ e $Y > t$. Portanto

$$\begin{aligned} F_V(t) &= P(V \leq t) = 1 - P(V > t) = 1 - P(X > t, Y > t) = 1 - P(X > t)P(Y > t) \\ &= 1 - (1 - F_X(t))(1 - F_Y(t)) = F_X(t) + F_Y(t) - F_X(t)F_Y(t). \end{aligned}$$

Nos pontos de continuidade de f_X e f_Y derivamos esta igualdade para obtermos

$$f_V(t) = f_X(t) + f_Y(t) - f_X(t)F_Y(t) - F_X(t)f_Y(t).$$

14.24. Exercícios

1. Sejam X e Y duas variáveis aleatórias independentes unidimensionais, cada uma com distribuição uniforme sobre o intervalo $[0, 1]$. Sejam $U = X + Y$ e $V = X - Y$.

(a) Provar que U tem uma função densidade contínua f_U dada por

$$f_U(u) = \begin{cases} u & \text{se } 0 < u \leq 1, \\ 2 - u & \text{se } 1 < u < 2, \\ 0 & \text{para os restantes valores de } u \end{cases}$$

(b) Definir, de modo análogo, uma densidade contínua f_V para V .

(c) Verificar se U e V são ou não independentes.

2. Sejam X e Y as variáveis aleatórias do Exercício 1, e $U = \max\{X, Y\}$, $V = \min\{X, Y\}$.

(a) Provar que U tem uma função densidade tal que $f_U(t) = 2t$ para $0 \leq t < 1$, e $f_U(t) = 0$ para os restantes valores de t .

(b) Definir a função densidade f_V para V .

(c) Verificar se U e V são ou não independentes.

3. Sejam X e Y variáveis aleatórias unidimensionais e independentes, cada um tendo uma distribuição exponencial de parâmetro $\lambda = 1$, e seja $f(x, y) = f_X(x)f_Y(y)$ o produto das densidades de X e Y .
- (a) Seja A o conjunto de pontos do plano xy para os quais $f(x, y) > 0$. Desenhar A e a sua image A' na aplicação definida por $u = x + y, v = x/(x + y)$.
- (b) Sejam $U = X + Y$ e $V = X/(X + Y)$ duas novas variáveis aleatórias. Calcular a densidade g de (U, V) .
- (c) Calcular a densidade f_U .
- (d) Calcular a densidade f_V .
4. Sejam X e Y duas variáveis aleatórias independentes, ambas com distribuição normal (media = 0, variância = 1). Introduzem-se novas variáveis aleatórias U e V definidas por $U = X/Y, V = Y$. Seja g a função densidade de (U, V) .
- (a) Mostrar que a função densidade de probabilidade de (U, V) é dada pela fórmula

$$g(u, v) = -\frac{v}{2\pi} e^{-(1+u^2)v^2/2} \quad \text{se } v < 0.$$

- (b) Determinar uma fórmula semelhante para calcular $g(u, v)$ quando $v \geq 0$.
- (c) Determinar uma função densidade de probabilidade de U .
5. Supor que X tem a densidade de probabilidade dada por

$$f_X(x) = \begin{cases} \frac{1}{\pi\sqrt{1-x^2}} & \text{se } -1 < x < 1, \\ 0 & \text{se } |x| \geq 1. \end{cases}$$

Se uma variável aleatória independente Y tem densidade

$$f_Y(y) = \begin{cases} ye^{-y^2/2} & \text{se } y \geq 0, \\ 0 & \text{se } y < 0, \end{cases}$$

determinar uma função densidade de $Z = XY$.

6. Dadas duas variáveis aleatórias unidimensionais independentes X e Y com densidades contínuas f_X e f_Y , sejam U e V duas novas variáveis aleatórias tais que $X = U \cos V, Y = U \sin V$, com $U > 0$ e $-\pi < V \leq \pi$.
- (a) Provar que U tem uma densidade tal que $f_U(u) = 0$ para $u < 0$, e

$$f_U(u) = u \int_{-\pi}^{\pi} f_X(u \cos v) f_Y(u \sin v) dv \quad \text{para } u \geq 0.$$

- (b) Determinar f_U e a correspondente função de repartição F_U quando cada uma das variáveis X e Y tem uma distribuição normal com média $m = 0$ e variância σ^2 .
7. (a) Supondo $\sigma_1 > 0$ e $\sigma_2 > 0$, verificar a identidade algébrica

$$\left(\frac{x - m_1}{\sigma_1}\right)^2 + \left(\frac{t - x - m_2}{\sigma_2}\right)^2 = \left(\frac{x - m_0}{\sigma_0}\right)^2 + \left(\frac{t - (m_1 + m_2)}{\sigma}\right)^2,$$

onde

$$\sigma^2 = \sigma_1^2 + \sigma_2^2, \quad \sigma_0^2 = \frac{\sigma_1^2 \sigma_2^2}{\sigma^2}, \quad \text{e} \quad m_0 = \frac{m_1 \sigma_2^2 + (t - m_2) \sigma_1^2}{\sigma_1^2 + \sigma_2^2}.$$

(b) Dadas duas variáveis aleatórias unidimensionais, independentes X e Y , supôr que X tem uma distribuição normal com média m_1 e variância σ_1^2 , e que Y tem uma distribuição normal com média m_2 e variância σ_2^2 . Provar que $X + Y$ tem uma distribuição normal com média $m = m_1 + m_2$ e variância $\sigma^2 = \sigma_1^2 + \sigma_2^2$.

8. Dadas duas variáveis aleatórias unidimensionais X e Y com densidades f_X e f_Y e densidade conjunta f . Para cada y fixo, define-se

$$f_X(x | Y = y) = \frac{f(x, y)}{f_Y(y)} \quad \text{sempre que } f_Y(y) > 0.$$

Esta é a chamada probabilidade condicionada de X , supondo que $Y = y$. De modo análogo define-se a densidade de probabilidade condicional de Y , supondo que $X = x$, pela fórmula

$$f_Y(y | X = x) = \frac{f(x, y)}{f_X(x)} \quad \text{sempre que } f_X(x) > 0.$$

- (a) Se f_Y e f_X são positivas, provar que $\int_{-\infty}^{\infty} f_X(x | Y = y) dx = \int_{-\infty}^{\infty} f_Y(y | X = x) dy = 1$.
 (b) Se f_Y e f_X são positivos, provar que

$$f_X(x) = \int_{-\infty}^{\infty} f_Y(y) f_X(x | Y = y) dy \quad \text{e} \quad f_Y(y) = \int_{-\infty}^{\infty} f_X(x) f_Y(y | X = x) dx.$$

9. Uma variável aleatória (X, Y) diz-se ter uma *distribuição normal bivariada* se a sua função densidade f é dada pela fórmula.

$$f(x, y) = \frac{\sqrt{D}}{2\pi} e^{-Q(x, y)/2},$$

onde $Q(x, y)$ é a forma quadrática

$$Q(x, y) = A_{11}(x - x_0)^2 + 2A_{12}(x - x_0)(y - y_0) + A_{22}(y - y_0)^2.$$

Os números A_{11} , A_{12} , A_{22} são constantes com $A_{11} > 0$. O número $D = A_{11}A_{22} - A_{12}^2$ chama-se o *discriminante* de Q e supõe-se positivo. Os números x_0 e y_0 são arbitrários.

- (a) Provar que $Q(x, y)$ pode exprimir-se como uma soma de quadrados como segue:

$$Q(x, y) = A_{11} \left(u + \frac{A_{12}}{A_{11}} v \right)^2 + \frac{D}{A_{11}} v^2, \quad \text{onde} \quad u = x - x_0, v = y - y_0.$$

- (b) Definir o integral duplo "impróprio" $\iint_{-\infty}^{+\infty} f(x, y) dx dy$ como o limite:

$$\iint_{-\infty}^{+\infty} f(x, y) dx dy = \lim_{t \rightarrow +\infty} \iint_{R(t)} f(x, y) dx dy,$$

onde $R(t)$ é o quadrado $[-t, t] \times [-t, t]$. Provar que

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1.$$

[Sugestão: Utilizar a alínea (a) para transformar o integral duplo sobre $R(t)$ num integral duplo no plano uv . Efectuar depois uma mudança de variáveis linear para simplificar o integral e, finalmente, fazer $t \rightarrow +\infty$.]

10. Se uma variável aleatória bidimensional (X, Y) tem uma distribuição normal bivariada como se indicou no Exercício 9, provar que X e Y são a si mesmas variáveis aleatórias unidimensionais normais com médias x_0 e y_0 , respectivamente, e com variâncias $\sigma^2(X) = A_{22}/D$, $\sigma^2(Y) = A_{11}/D$.
11. Se (X, Y) tem uma distribuição normal bivariada como se indica no Exercício 9, provar que a variável aleatória $Z = X + Y$ tem uma distribuição normal unidimensional com média $x_0 + y_0$ e variância $(A_{11} - 2A_{12} + A_{22})/D$.

14.25 Esperança matemática e variância

A interpretação, em termos de massa, das distribuições de probabilidade pode levar-se a um estágio mais avançado introduzindo os conceitos de *esperança matemática* e *variância*. Estes desempenham na teoria das probabilidades o meso papel que o “centro de massa” e o “momento de inércia” desempenham na mecânica. Não utilizando o integral de Stieltjes temos que apresentar separadamente definições para os casos discreto e contínuo.

DEFINIÇÕES DE ESPERANÇA MATEMÁTICA E VARIÂNCIA. Se X é uma variável aleatória unidimensional, a *esperança matemática* e a *variância* X são números reais representados por $E(X)$ e $\text{Var}(X)$ respectivamente, e definem-se do modo seguinte:

(a) Para uma variável aleatória com função densidade f .

$$E(X) = \int_{-\infty}^{+\infty} t f_X(t) dt,$$

$$\text{Var}(X) = \int_{-\infty}^{+\infty} [t - E(X)]^2 f_X(t) dt.$$

(b) Para uma variável aleatória discreta com pontos de massa x_1, x_2, \dots tendo as probabilidades $p_k = P(X = x_k)$

$$E(X) = \sum_{k=1}^{\infty} x_k p_k,$$

$$\text{Var}(X) = \sum_{k=1}^{\infty} [x_k - E(X)]^2 p_k.$$

Nota: $E(X)$ e $\text{Var}(X)$ existem unicamente quando o integral ou as séries são absolutamente convergentes. Subentende-se que a série é uma soma finita quando o conjunto fundamental for finito; em tal hipótese $E(X)$ e $\text{Var}(X)$ existem sempre. Também existem quando f_X é 0 no exterior de algum intervalo finito.

A esperança matemática $E(X)$ é um valor teoricamente calculado associado à variável aleatória X . Em certas circunstâncias, a distribuição actua como se toda a massa estivesse concentrada num único ponto, $E(X)$. O verdadeiro significado da esperança matemática na teoria das probabilidades será analisado na Secção 14.29 em conexão com a chamada “lei dos grandes números”.

Em mecânica, o conhecimento do centro de massa por si só não dá qualquer indicação acerca do modo como a massa se dispersa em relação ao seu centro. Uma medida dessa dispersão é garantida pelo “momento de segunda ordem” ou “momento de inércia”. Na teoria das probabilidades, este momento de segunda ordem é a variância. Esta mede a tendência de uma distribuição em dispersar-se relativaente à esperança matemática. Na Secção 14.28 verificaremos que um valor pequeno da variância significa que são improváveis os grandes desvios relativamente ao valor da esperança matemática.

Embora a esperança matemática $E(X)$ possa ser positiva ou negativa, a variância $\text{Var}(X)$ é sempre não negativa. O símbolo σ^2 é também utilizado para representar a variância. A sua raiz quadrada positiva chama-se o *desvio padrão*, *desvio tipo* ou *desvio quadrático médio* e representa-se por σ . O desvio padrão é uma média pesada; o conceito análogo na mecânica é o “raio de giração”.

EXEMPLO 1. Distribuição uniforme. Seja X uma variável aleatória com distribuição uniforme sobre um intervalo $[a, b]$. Então $f(t) = 1/(b - a)$ se $a < t < b$, e $f(t) = 0$ para os restantes valores. Deste modo a esperança matemática de X é dada por

$$E(X) = \int_{-\infty}^{+\infty} tf(t) dt = \frac{1}{b-a} \int_a^b t dt = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}.$$

Assim a média com esperança matemática, é o ponto médio do intervalo. Se fizermos $(a+b)/2 = m$ e consideramos que $m - a = b - m = (b-a)/2$, encontramos

$$\text{Var}(X) = \frac{1}{b-a} \int_a^b (t-m)^2 dt = \frac{1}{b-a} \int_{a-m}^{b-m} u^2 du = \frac{(b-a)^2}{12}.$$

Observe-se que a variância depende unicamente do *comprimento* do intervalo.

EXEMPLO 2. Distribuição binomial. Se X tem uma distribuição binomial com parâmetros n e p temos

$$E(X) = \sum_{k=0}^n k \binom{n}{k} p^k q^{n-k},$$

com $q = 1 - p$. Para calcular esta soma, consideramos

$$f(x, y) = (x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}$$

e observamos que

$$\sum_{k=0}^n k \binom{n}{k} x^{k-1} y^{n-k} = \frac{\partial f(x, y)}{\partial x} = n(x + y)^{n-1}.$$

Se multiplicamos ambos os membros por x e fazemos $x = p$ e $y = q$, obtemos $E(X) = np$.

Por um raciocínio análogo podemos deduzir a fórmula

$$\text{Var}(X) = \sum_{k=0}^n (k - np)^2 \binom{n}{k} p^k q^{n-k} = npq.$$

No exercício 6 da Secção 14.27 sugerem-se duas demonstrações desta fórmula.

EXEMPLO 3. Distribuição normal. Os termos “média” e “variância” já foram introduzidas a quando do estudo da distribuição normal na Secção 14.14. Tais designações justificam-se pelas fórmulas

$$E(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} te^{-[t-m]^2/\sigma^2} dt = m$$

$$\text{Var}(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (t - m)^2 e^{-[t-m]^2/\sigma^2} dt = \sigma^2.$$

No Exercício 7 da Secção 14.27 são pedidas as demonstrações destas fórmulas.

Os jogadores recorrem frequentemente ao conceito de esperança matemática para decidirem se determinado jogo de azar lhes é ou não favorável. Como exemplo vamos considerar o jogo de aposta no “vermelho” ou “preto” na roleta.

EXEMPLO 4. Roleta. Uma roleta percorre os números de 0 a 36. O número 0 aparece sobre fundo cinzento, metade dos 36 números restantes sobre fundo vermelho, e a outra metade sobre fundo preto. Os métodos correntes de apostas são:

- (1) Apostar 1 \$00 sobre uma cor (vermelho ou preto). Ganho possível: 2 \$00.
- (2) Apostar 1 \$00 num só número (0 excluído). Ganho possível 36 \$00.
- (3) Apostar 1 \$00 numa dúzia qualquer (0 excluído). Ganho possível 3 \$00.

Se sai o zero a branca ganha e todos os restantes jogadores perdem.

Seja X a variável aleatória que mede o ganho do jogador da aposta na hipótese (1). Os valores possíveis de X são $x_1 = -1$ e $x_2 = +1$. As probabilidades são $P(X = x_1) = \frac{19}{37}$, $P(X = x_2) = \frac{18}{37}$. Portanto a esperança matemática é

$$E(X) = (-1)\frac{19}{37} + (+1)\frac{18}{37} = -\frac{1}{37};$$

este resultado é usualmente interpretado como significado que o jogo é desfavorável a quem o jogue. A justificação matemática desta interpretação é feita por uma das leis dos grandes números, que estudaremos na Secção 14.29. O leitor pode verificar que a esperança matemática tem o mesmo valor para as apostas das hipóteses (2) e (3).

EXEMPLO 5. Jogo de lançamento da moeda. Num jogo de lançamento de uma moeda existe uma probabilidade p que saia a face “cara” (H) e uma probabilidade q que saia a face “cruz” (T), com $0 \leq p \leq 1$ e $q = 1 - p$. Lança-se a moeda ao ar repetidas vezes até que ocorra de novo o resultado do primeiro lançamento; uma vez conseguido isso termina o jogo. Se no primeiro lançamento tiver saído “cara” (H) o jogador recebe 1 \$000 por cada T que saia a seguir até voltar a sair H . Por exemplo, se a sequência de resultado é $HTTTH$ o jogador recebe 3\$00, mas se a sequência for HH não recebe nada. Se o primeiro resultado for T aplica-se a mesma regra de jogo trocando H por T . O problema consiste em determinar quanto deverá pagar inicialmente o jogador para participar no jogo. Para o determinarmos consideramos a variável aleatória que representa o número de escudos ganhos e calculamos a sua esperança matemática.

Para o conjunto fundamental tomamos a colecção de todas as partidas possíveis que possam ocorrer. Tal conjunto pode exprimir-se como a reunião de dois conjuntos A e B , em que

$$A = \{TT, THT, THHT, THHHT, \dots\} \quad \text{e} \quad B = \{HH, HTH, HTTH, HTTTH, \dots\}.$$

Os elementos do conjunto A (pela ordem indicada) designamo-los por a_0, a_1, a_2, \dots e os do conjunto B por b_0, b_1, b_2, \dots . Em seguida definimos as probabilidades do modo seguinte:

$$P(a_n) = p^n q^2 \quad \text{e} \quad P(b_n) = q^n p^2.$$

(Quando $p = 0$, fazemos $P(a_0) = 1$ e $P(x) = 0$ para qualquer outro x de $A \cup B$. Quando $q = 0$, fazemos $P(b_0) = 1$ e $P(x) = 0$ para os restantes x). Na Secção 13.21 provar-se-á que esta distribuição de probabilidades é aceitável.

A variável aleatória X que nos interessa define-se sobre o conjunto fundamental $A \cup B$ do modo seguinte:

$$X(a_n) = X(b_n) = n \quad \text{para } n = 0, 1, 2, \dots$$

O acontecimento “ $X = n$ ” consiste das duas partidas a_n e b_n , pelo que se tem

$$P(X = n) = p^n q^2 + q^n p^2,$$

onde p^0 e q^0 se consideram com o valor 1 quando $p = 0$ ou $q = 0$. A esperança matemática de X é dada pela soma

$$(14.40) \quad E(X) = \sum_{n=0}^{\infty} n P(X = n) = q^2 \sum_{n=0}^{\infty} n p^n + p^2 \sum_{n=0}^{\infty} n q^n.$$

Se fôr $p = 0$ ou $q = 0$, obtemos $E(X) = 0$. Por outro lado podemos calcular as somas das séries em (14.40) considerando que para $0 < x < 1$ se tem

$$\sum_{n=0}^{\infty} nx^n = x \frac{d}{dx} \sum_{n=0}^{\infty} x^n = x \frac{d}{dx} \left(\frac{1}{1-x} \right) = \frac{x}{(1-x)^2}.$$

Usando esta fórmula em (14.40) com $x = p$ e $x = q$ obtemos, para $0 < p < 1$,

$$E(X) = \frac{q^2 p}{(1-p)^2} + \frac{p^2 q}{(1-q)^2} = p + q = 1.$$

Interpreta-se este resultado dizendo que o jogo é desfavorável para o jogador que pague mais que 1 \$ 00 para participar nele.

Este exemplo particular é de especial interesse porque a esperança matemática $E(X)$ é independente de p quando $0 < p < 1$. Por outras palavras, viciando a moeda a favor da saída da face "cara" ou da face "cruz", tal não afecta o valor da esperança matemática excepto nos casos extremos em que a moeda seja de tal modo viciada que nunca saia a face "cara" ou a face "cruz". Observe-se que, como função de p , a esperança matemática $E(X)$ é descontínua nos pontos $p = 0$ e $p = 1$. Nos casos restantes tem o valor constante 1. Este interessante exemplo foi sugerido ao autor por H. S. Zuckermann.

14.26 Esperança matemática de uma função de uma variável aleatória

Se uma nova variável aleatória Y está relacionada com uma dada variável X por uma equação da forma $Y = \varphi(X)$, a sua esperança matemática define-se (no caso contínuo) pela fórmula

$$(14.41) \quad E(Y) = \int_{-\infty}^{+\infty} t f_Y(t) dt.$$

A esperança matemática $E(Y)$ pode calcular-se directamente em função da densidade f_X sem determinar a densidade de Y . Com efeito, a fórmula seguinte é equivalente a (14.41)

$$(14.42) \quad E(Y) = \int_{-\infty}^{+\infty} \varphi(t) f_X(t) dt.$$

Uma demonstração de (14.42) no caso mais geral apresenta alguma dificuldade e não será tentada aqui. Todavia, para muitos casos particulares importantes a demonstração é simples. Nestes casos φ é derivável e estritamente crescente em todo o eixo real. Para uma variável aleatória X , correspondente a uma distribuição contínua, com função densidade f_X temos a seguinte fórmula que define a densidade de Y (estabelecida na Secção 14.17):

$$f_Y(t) = f_X[\varphi(t)] \cdot \varphi'(t),$$

com $\psi(t)$ a função inversa de $\varphi(t)$. Substituindo em (14.41) e efectuando a mudança de variável $u = \psi(t)$ [pelo que $t = \varphi(u)$] obtemos

$$E(Y) = \int_{-\infty}^{+\infty} t f_Y(t) dt = \int_{-\infty}^{+\infty} t f_X[\psi(t)] \cdot \psi'(t) dt = \int_{-\infty}^{+\infty} \varphi(u) f_X(u) du,$$

a qual é a mesma que (14.42).

Quando (14.42) se aplica a $Y = (X - m)^2$ onde $m = E(X)$, obtemos

$$E(Y) = \int_{-\infty}^{+\infty} (t - m)^2 f_X(t) dt = \text{Var}(X).$$

Prova esta igualdade que a variância é ela própria uma esperança matemática. Também no caso discreto é, naturalmente, válida uma fórmula análoga a (14.42). Mais geralmente, pode provar-se que

$$E[\varphi(X, Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \varphi(x, y) f(x, y) dx dy$$

se (X, Y) for uma variável aleatória contínua com densidade conjunta f .

Nota: Para variáveis aleatórias bidimensionais, a esperança matemática e a variância podem definir-se de maneira análoga à do caso unidimensional, excepto que passam a utilizar-se integrais duplos e duplas somas. Não analisaremos aqui tal extensão.

14.27 Exercícios

1. Lança-se um dado. Seja X o número de pontos obtido. Calcular $E(X)$ e $\text{Var}(X)$.
2. Admitir que X é uma variável aleatória com uma função densidade de probabilidade. Seja $Y = (X - m)/\sigma$, onde $m = E(X)$ e $\sigma = \sqrt{\text{Var}(X)}$. Provar que $E(Y) = 0$ e $E(Y^2) = 1$.
3. Estabelecer as seguintes propriedades gerais da esperança matemática e da variância, quer para o caso discreto, quer para o caso contínuo.
 - (a) $E(cX) = cE(X)$, com c constante.
 - (b) $\text{Var}(cX) = c^2 \text{Var}(X)$, com c constante.
 - (c) $E(X + Y) = E(X) + E(Y)$.
 - (d) $\text{Var}(X) = E(X^2) - [E(X)]^2$.
 - (e) $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2E[(X - E(X))(Y - E(Y))]$.
 - (f) $E[\varphi_1(X) + \varphi_2(Y)] = E[\varphi_1(X)] + E[\varphi_2(Y)]$. [A alínea (c) é um caso particular.]
4. Se X e Y são variáveis aleatórias independentes, provar que
 - (a) $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.
 - (b) $E[\varphi(X) \cdot \psi(Y)] = E[\varphi(X)] \cdot E[\psi(Y)]$.
 - (c) Se X_1, X_2, \dots, X_n são variáveis aleatórias independentes com $E(X_k) = m_k$, provar que

$$\text{Var} \left[\sum_{k=1}^n (X_k - m_k) \right] = \sum_{k=1}^n \text{Var}(X_k - m_k) = \sum_{k=1}^n \text{Var}(X_k).$$

5. Sejam $X_1, X_2, X_3, \dots, X_n$, n variáveis aleatórias independentes, com a mesma esperança matemática, $E(X_k) = m$, e a mesma variância, $\text{Var}(X_k) = \sigma^2$. Represente \bar{X} a média aritmética, $\bar{X} = (1/n) \sum_{i=1}^n X_i$. Utilizar os Exercícios 3 e 4 para provar que $E(\bar{X}) = m$ e $\text{Var}(\bar{X}) = \sigma^2/n$.
6. (a) Se $q = 1 - p$, demonstrar a fórmula

$$\sum_{k=0}^n (k - np)^2 \binom{n}{k} p^k q^{n-k} = npq,$$

provando desse modo que $\text{Var}(X) = npq$ para uma variável aleatória X tendo uma distribuição binomial com parâmetros n e p . [Sugestão: $k^2 = k(k-1) + k$].

(b) Se X tem uma distribuição binomial com parâmetros n e p , provar que X pode exprimir-se como uma soma de n variáveis aleatórias independentes X_1, X_2, \dots, X_n , cada uma delas podendo tomar os valores possíveis 0 e 1 com probabilidades p e q , respectivamente, e tendo cada uma uma distribuição binomial. Utilizar este resultado e o Exercício 5 para provar que $E(X) = np$ e $\text{Var}(X) = npq$.

7. Determinar a esperança matemática e a variância (sempre que existam) para uma variável aleatória X tendo

- (a) uma distribuição de Poisson com parâmetro λ .
- (b) uma distribuição de Cauchy.
- (c) uma distribuição exponencial com parâmetro λ .
- (d) uma distribuição normal.

8. Uma variável aleatória X tem uma função densidade de probabilidade definida por

$$f(t) = \frac{C(r)}{|t|^r} \quad \text{se } |t| > 1, \quad f(t) = 0 \quad \text{se } |t| \leq 1,$$

onde $r > 1$ e $C(r)$ é independente de t .

(a) Exprimir $C(r)$ em função de r e fazer um desenho que dê ideia da natureza do gráfico de f .

(b) Determinar a correspondente função de repartição F_X e fazer um desenho que dê ideia do seu gráfico.

(c) Calcular $P(X < 5)$ e $P(5 < X < 10)$ em função de r .

(d) Para que valores de r terá X uma esperança matemática finita? Calcular $E(X)$ em função de r quando a esperança matemática for finita.

(e) Para que valores de r terá X uma variância finita? Calcular $E(X)$ em função de r , quando for finita.

9. Um jogador de roleta joga segundo o seguinte "sistema". Joga um conjunto de três partidas. Na primeira e segundas partidas aposta 10\$00 no vermelho. Na terceira partida procede do seguinte modo:

- (a) Se ganhou nas duas primeiras partidas, não aposta.
- (b) Se ganhou numa das duas primeiras partidas e perdeu na outra, aposta 10\$00 na cor "oposto" à que saiu na segunda partida.
- (c) Se perdeu em duas primeiras partidas, aposta 30\$00 no vermelho.

Representem X , Y , e Z , respectivamente, os resultados dos lucros na primeira, segunda e terceira partidas. Calcular $E(X)$, $E(Y)$, $E(Z)$ e $E(X + Y + Z)$.

10. (*Problema de Petersburgo*). Um jogador lança uma moeda e ganha 1 \$00 se nesse lançamento sai "cara". Se sai outra vez "cara" ganha outra vez 1 \$00. Se sai "cara" uma terceira vez ganha 2 \$00 (no total ganha 4 \$00). Desde que em sucessivos lançamentos obtenha "cara" n vezes acumula um ganho de 2^{n-1} escudos. O jogo termina quando sair a face "cruz". Seja X o número de escudos ganhos em determinado jogo. Calcular $E(X)$. Em face do resultado quanto deverá um jogador pagar de entrada para participar no jogo nas condições indicadas?
11. (a) Supor X uma variável aleatória contínua com densidade de probabilidade f_X . Seja $Y = (X - m)/\sigma$, onde $m = E(X)$ e $\sigma = \sqrt{\text{Var}(X)}$. Provar que

$$E(e^Y) = e^{-m/\sigma} \int_{-\infty}^{+\infty} e^{t/\sigma} f_X(t) dt.$$

- (b) Seja X uma variável aleatória discreta tendo uma distribuição de Poisson com parâmetro λ . Definir Y como na alínea (a) e provar que

$$E(e^Y) = e^{-\lambda G(\lambda)}, \quad \text{com} \quad G(\lambda) = 1 + \frac{1}{\sqrt{\lambda}} - e^{1/\sqrt{\lambda}}.$$

12. Uma variável aleatória X tem uma distribuição normal standard. Calcular: (a) $E(|X|)$; (b) $E(e^X)$; (c) $\text{Var}(e^X)$; (d) $E(\sqrt{X^2 + Y^2})$. Na alínea (d), Y tem também uma distribuição normal standard, mas é independente de X .

14.28. Desigualdade de Tchebycheff

Como atrás já referimos, um pequeno valor da variância significa que é improvável que uma variável aleatória X se desvie muito da sua esperança matemática ou média. Para tornarmos mais precisa esta afirmação introduzimos o valor absoluto $|X - E(X)|$ que mede a distância entre X e $E(X)$. Qual a probabilidade de que esta distância seja maior do que um dado valor? Para obter resposta para esta pergunta devemos determinar a probabilidade

$$P[|X - E(X)| > c],$$

com c um número positivo dado. No caso contínuo temos

$$\begin{aligned} P[|X - E(X)| > c] &= 1 - P[|X - E(X)| \leq c] = 1 - P[E(X) - c \leq X \leq E(X) + c] \\ &= \int_{-\infty}^{+\infty} f_X(t) dt - \int_{E(X)-c}^{E(X)+c} f_X(t) dt \\ (14.43) \quad &= \int_{-\infty}^{E(X)-c} f_X(t) dt + \int_{E(X)+c}^{+\infty} f_X(t) dt; \end{aligned}$$

portanto o cálculo dessa probabilidade pode efectuar-se desde que seja conhecida a função densidade f_X . Naturalmente se f_X é desconhecida este método não nos dá qualquer informação. Todavia se a *variância* é conhecida, podemos obter um limite supe-

rior para esta probabilidade. Este limite superior é determinado pelo teorema que a seguir se apresenta devido a P. L. Tchebycheff (1821-1894), um famoso matemático russo que deu notável contribuição ao desenvolvimento da teoria das probabilidades e de outros ramos da matemática, em especial a teoria dos números.

TEOREMA 14.11. DESIGUALDADE DE TCHEBYCHEFF. *Se X é uma variável aleatória unidimensional com esperança matemática $E(X)$ finita e variância $\text{Var}(X)$, então para todo o número positivo c tem-se*

$$(14.44) \quad P[|X - E(X)| > c] \leq \frac{\text{Var}(X)}{c^2}.$$

Demonstração. No caso contínuo temos

$$\begin{aligned} \text{Var}(X) &= \int_{-\infty}^{+\infty} [t - E(X)]^2 f_X(t) dt \\ &\geq \int_{-\infty}^{E(X)-c} [t - E(X)]^2 f_X(t) dt + \int_{E(X)+c}^{+\infty} [t - E(X)]^2 f_X(t) dt \\ &\geq c^2 \left(\int_{-\infty}^{E(X)-c} f_X(t) dt + \int_{E(X)+c}^{+\infty} f_X(t) dt \right). \end{aligned}$$

Devido a (14.43), o coeficiente de c^2 no segundo membro é $P[|X - E(X)| > c]$. Portanto, dividindo por c^2 ambos os membros, obtemos (14.44). Está pois demonstrado o teorema para o caso contínuo; o caso discreto pode tratar-se de maneira semelhante.

A desigualdade de Tchebycheff diz-nos que quanto maior for c menor será a probabilidade de que $|X - E(X)| > c$. Por outras palavras, é improvável que X se desvie muito de $E(X)$; tal facto é ainda mais improvável se a variância $\text{Var}(X)$ for pequena.

Se substituirmos c por $k\sigma$, onde $k > 0$ e σ representa o desvio padrão, $\sigma = \sqrt{\text{Var}(X)}$, a desigualdade de Tchebycheff escreve-se

$$P[|X - E(X)| > k\sigma] \leq \frac{1}{k^2}.$$

Quer dizer, a probabilidade de que X difira da sua esperança matemática por um valor superior a k vezes o seu desvio padrão não excede $1/k^2$. Por exemplo, quando $k = 10$ esta desigualdade diz-nos que a probabilidade $P[|X - E(X)| > 10\sigma]$ não excede 0,010. Por outras palavras, a probabilidade de que um valor observado de X difira da esperança matemática mais do que dez vezes o desvio padrão é não superior a 0,010. Analogamente, quando $k = 3$ encontramos que a probabilidade de que um valor observado defiria da média em mais do que três vezes o desvio padrão não excede 0,111.

A desigualdade de Tchebycheff é um teorema geral que se aplica a todas as distribuições. Em muitas aplicações a desigualdade pode tornar-se mais forte quando se tem mais informação acerca da distribuição que se considera. Por exemplo, se X tem uma distribuição binomial com parâmetros n e p pode provar-se (recorrendo à apro-

ximação normal da distribuição binomial) que para grandes valores de n é cerca de 0,003 a probabilidade de que um valor observado defira da média por mais do que três vezes o desvio padrão. (Para este resultado, basta $n \geq 12$). Esta probabilidade é muito menor do que a probabilidade 0,111 determinada pela desigualdade de Tchebycheff.

EXEMPLO. *Verificação da "imparcialidade" de uma moeda.* Pretendemos verificar se sim ou não uma dada moeda é imparcial lançando-a ao ar 10.000 vezes e anotando o número de vezes em que ocorreu face "cara". Para uma moeda imparcial a variável aleatória X que dá conta do número de ocorrências face "cara" tem uma distribuição binomial com parâmetros $n = 10.000$ e $p = \frac{1}{2}$. A média de X é $np = 5.000$ e o desvio padrão é $\sigma = \sqrt{npq} = 50$. (Ver o Exemplo 2 da Secção 14.25). Como mencionámos anteriormente, a probabilidade de que uma variável aleatória com distribuição binomial defira da sua média por um valor superior a 3σ é cerca de 0,003. Deste modo diremos que uma moeda *não é imparcial* se o número de ocorrências da face "cara" em 10.000 lançamentos diferir da média por um valor superior a 3σ . Visto que $E(X) = 5.000$ e $3\sigma = 150$, poderíamos dizer que a moeda não é imparcial se o número de ocorrências da face "cara" em 10.000 lançamentos for menor que 4.850 ou maior que 5.150.

14.29. Leis dos grandes números

Em ligação com os problemas de lançamentos ao ar de moedas, diz-se muitas vezes que a probabilidade de obter face "cara" um lançamento de uma moeda imparcial é $\frac{1}{2}$. Não significa isto que se a moeda for lançada duas vezes necessariamente ocorrerá face "cara" precisamente uma vez. Nem tão pouco significa que em 1000 lançamentos ocorrerá face "cara" 500 vezes. Representemos por $h(n)$ o número de vezes que ocorre face "cara" em n lançamentos. Mostra a experiência que mesmo para n suficientemente grande, o cociente $h(n)/n$ não é necessariamente $\frac{1}{2}$. Contudo, mostra também a experiência que este cociente parece *aproximar-se de* $\frac{1}{2}$ quando n aumenta, embora durante o processo possa oscilar consideravelmente para mais ou para menos do valor $\frac{1}{2}$. Tal facto sugere-se que será possível provar que

$$(14.45) \quad \lim_{n \rightarrow \infty} \frac{h(n)}{n} = \frac{1}{2}.$$

Infelizmente tal não é possível. Uma dificuldade é que o número $h(n)$ depende não somente de n , mas também da experiência particular que está sendo realizada. Nós não temos maneira de saber de antemão como é que $h(n)$ variará de uma experiência para a outra. Mas a real dificuldade está em que é possível (embora não seja muito provável) que uma determinada experiência o cociente $h(n)/n$ possa *não* tender de nenhum modo para $\frac{1}{2}$. Por exemplo, não há razão para excluir a possibilidade de obtenção da face "cara" em *todos* os lançamentos da moeda, caso em que $h(n) = n$ e $h(n)/n \rightarrow 1$. Portanto, em vez de tentar provar (14.45), verificaremos ser mais razoável (e mais útil)

averiguar quanto será provável que $h(n)/n$ difira de $\frac{1}{2}$ numa certa quantidade. Quer dizer, dado um certo número positivo c , determinamos a probabilidade

$$P\left(\left|\frac{h(n)}{n} - \frac{1}{2}\right| > c\right).$$

Pela introdução de uma variável aleatória adequada e recorrendo à desigualdade de Tchebycheff podemos obter um *limite superior* útil para esta probabilidade, limite superior esse que não exige um conhecimento explícito de $h(n)$. Isto leva-nos a uma nova realação limite que substituirá (14.45) de forma adequada.

Não é necessário nenhum esforço suplementar para tratar o caso mais geral de experiências num esquema de Bernoulli, no qual a probabilidade do “acontecimento favorável” é p e a probabilidade do “acontecimento contrário” é q . (No lançamento de moedas ao ar “acontecimento favorável” pode significar saída da face “cara” e para p tomamos o valor $\frac{1}{2}$). Seja X a variável aleatória definida pelo número de acontecimentos favoráveis em n experiências independentes. Então X tem uma distribuição binomial com esperança matemática $E(X) = np$ e variância $\text{Var}(X) = npq$. Logo é aplicável a desigualdade de Tchebycheff, a qual estabelece que

$$(14.46) \quad P(|X - np| > c) \leq \frac{npq}{c^2}.$$

Visto que nos interessa o cociente X/n , que pode designar-se por *frequência relativa* do acontecimento favorável, dividimos a desigualdade $|X - np| > c$ por n e escrevemos de nova (14.46) na forma seguinte

$$(14.47) \quad P\left(\left|\frac{X}{n} - p\right| > \frac{c}{n}\right) \leq \frac{npq}{c^2}.$$

Porque (14.47) é válida para todo $c > 0$, podemos fazer c depender de n e escrever $c = \epsilon n$, com ϵ um número positivo fixo. Então (14.47) vem

$$P\left(\left|\frac{X}{n} - p\right| > \epsilon\right) \leq \frac{pq}{n\epsilon^2}.$$

O aparecimento de n no denominador do segundo membro sugere que façamos $n \rightarrow \infty$. somos assim conduzidos à formula limite

$$(14.48) \quad \lim_{n \rightarrow \infty} P\left(\left|\frac{X}{n} - p\right| > \epsilon\right) = 0 \quad \text{para todo } \epsilon > 0 \text{ fixo,}$$

chamada a *lei dos grandes números para a distribuição de Bernoulli*. Diz-nos esta lei que, dado um $\epsilon > 0$ (por pequeno que seja), a probabilidade de que a frequência rela-

tiva do acontecimento favorável defira de p por uma quantidade superior a ϵ é uma função de n que tende para zero quando $n \rightarrow \infty$. Esta relação limite dá-nos uma justificação matemática para a atribuição da probabilidade $\frac{1}{2}$ na obtenção de qualquer das faces num lançamento de uma moeda perfeita ao ar.

A relação limite (14.48) é um caso particular de um resultado mais geral no qual a "frequência relativa" X/n é substituída pela média aritmética de n variáveis aleatórias independentes tendo a mesma esperança matemática e a mesma variância. Este teorema mais geral é habitualmente designado por *lei fraca dos grandes números*; pode ser enunciada do modo seguinte:

TEOREMA 14.12. FORMA FRACA DA LEI DOS GRANDES NÚMEROS. *Sejam X_1, X_2, \dots, X_n , n variáveis aleatórias independentes, tendo todas a mesma média e a mesma variância a saber*

$$E(X_k) = m \quad \text{e} \quad \text{Var}(X_k) = \sigma^2 \quad \text{para} \quad k = 1, 2, \dots, n.$$

Define-se uma variável aleatória \bar{X} (chamada a média aritmética de X_1, X_2, \dots, X_n) pela fórmula

$$\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k.$$

Então, para cada $\epsilon > 0$, fixo, tem-se

$$(14.49) \quad \lim_{n \rightarrow \infty} P(|\bar{X} - m| > \epsilon) = 0.$$

ou, o que é a mesma coisa,

$$(14.50) \quad \lim_{n \rightarrow \infty} P(|\bar{X} - m| \leq \epsilon) = 1.$$

Demonstração. Aplicamos a desigualdade de Tchebycheff a \bar{X} . Para tal necessitamos conhecer a esperança matemática e a variância de \bar{X} . Estas são

$$E(\bar{X}) = m \quad \text{e} \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n}.$$

(Ver o Exercício 5 da Secção 14.27). A desigualdade de Tchebycheff escreve-se então $P(|\bar{X} - m| > c) \leq \sigma^2/(nc^2)$. Fazendo $n \rightarrow \infty$ e substituindo c por ϵ obtemos (14.49) e portanto (14.50).

Nota: Para provar que a relação limite (14.48) é um caso particular do teorema 14.12, supomos que cada X_k tem os valores possíveis 0 e 1, com probabilidades $P(X_k = 1) = p$ e $P(X_k = 0) = 1 - p$. Então \bar{X} é a frequência relativa do acontecimento favorável em n experiências independentes, $E(\bar{X}) = p$, e (14.49) reduz-se a (14.48).

O teorema 14.12 diz-se uma forma *fraca* da lei dos grandes números, porque existe também uma forma *forte* dessa mesma lei a qual (sob as mesmas hipóteses) estabelece que

$$(14.51) \quad P\left(\lim_{n \rightarrow \infty} |\bar{X} - m| = 0\right) = 1.$$

A principal diferença entre (14.51) e (14.50) é que as operações “limite” e “probabilidade” estão permutados. Pode provar-se que a forma forte implica a forma fraca, mas a inversa não é verdadeira.

Observe-se que a forma forte em (14.51) parece mais próxima de (14.45) que (14.50). Com efeito, (14.51) diz que se tem $\lim_{n \rightarrow \infty} \bar{X} = m$ “quase sempre”, isto é, com probabilidade 1. Quando aplicada ao lançamento da moeda, em particular, diz-nos que a não verificação de (14.45) não é mais provável do que obter sempre cara no lançamento repetido de uma moeda. A forma forte mostra realmente porquê a teoria das probabilidades corresponde à experiência e à nossa sensação intuitiva do que “deve ser” a probabilidade.

A demonstração da forma forte da lei dos grandes número é muito extensa e por tal razão omiti-la-emos. Nos livros indicados com os números 1, 3, 8 e 10 na Bibliografia do final deste capítulo aparecem demonstrações deste teorema.

14.30. Teorema limite central do cálculo das probabilidades

Em muitas aplicações da teoria das probabilidades, intervêm variáveis aleatórias que são somas de outras variáveis aleatórias. Por exemplo, o lucro depois de várias partidas de um jogo é a soma dos ganhos em cada um deles. Um facto surpreendente se verifica quando um grande número de variáveis aleatórias se adicionam. Sob certas condições gerais (aplicáveis em quase todas as situações que ocorrem na prática) a distribuição da soma tende a ser normal, independentemente da distribuição de cada uma das aleatórias parcelas que intervêm nessa soma. A proposição correcta deste facto notável é conhecida por *teorema limite central do cálculo das probabilidades* e justifica a importância da distribuição normal, quer na teoria, quer na prática. A discussão exhaustiva deste teorema pertence a estudos mais avançados da teoria das probabilidades. Aqui limitar-nos-emos a explicar o que o teorema afirma.

Suponhamos uma sucessão infinita de variáveis aleatórias, por exemplo X_1, X_2, \dots cada uma com esperança matemática e variância finita. Seja

$$m_k = E(X_k) \quad \text{e} \quad \sigma_k^2 = \text{Var}(X_k), \quad k = 1, 2, \dots$$

Construimos uma nova variável aleatória S_n adicionando as n primeiras diferenças $X_k - m_k$:

$$(14.52) \quad S_n = \sum_{k=1}^n (X_k - m_k).$$

Adicionando as diferenças em vez das próprias variáveis aleatórias X_k , então a soma S_n terá uma esperança matemática 0. O problema consiste em determinar o limite da função de repartição de S_n quando $n \rightarrow \infty$.

Se X_1, X_2, \dots, X_n são independentes, então [pelo Exercício 4(c) da Secção 14.27] temos

$$\text{Var}(S_n) = \sum_{k=1}^n \text{Var}(X_k - m_k) = \sum_{k=1}^n \text{Var}(X_k) = \sum_{k=1}^n \sigma_k^2.$$

Habitualmente $\text{Var}(S_n)$ será grande, muito embora as variâncias individuais σ_k^2 possam ser pequenas. Variáveis aleatórias com grandes valores para a variância não interessam porque os seus valores tendem a apresentar uma grande dispersão relativamente à média. Por tal motivo introduz-se uma nova variável aleatória T_n definida por

$$(14.53) \quad T_n = \frac{S_n}{\sqrt{\text{Var}(S_n)}}.$$

Esta nova variável tem média 0 e variância 1 e diz-se uma variável aleatória *reduzida*. A variável aleatória reduzida T_n tem significado, mesmo ainda se as variáveis aleatórias X_1, X_2, \dots, X_n não são independentes.

Introduzimos agora a seguinte definição:

DEFINIÇÃO DA PROPRIEDADE LIMITE CENTRAL. *Seja*

$$(14.54) \quad X_1, X_2, X_3, \dots$$

uma sucessão de variáveis aleatórias (não necessariamente independentes), onde cada X_k tem uma esperança matemática m_k e uma variância σ_k^2 , ambas finitas. Definem-se S_n e T_n por (14.52) e (14.53). A sucessão (14.54) diz-se satisfazer à propriedade limite central se, para quaisquer a e b com $a \leq b$, se tem

$$(14.55) \quad \lim_{n \rightarrow \infty} P(a \leq T_n \leq b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-u^2/2} du.$$

Por outras palavras, as variáveis aleatórias em (14.54) satisfazem à propriedade limite central se a variável aleatória T_n tem uma distribuição que tende para a distribuição da variável normal reduzida quando $n \rightarrow \infty$. [A equação (14.55) é válida também se $a = -\infty$ ou $b = +\infty$.]

Laplace foi o primeiro a verificar que esta propriedade é satisfeita por muitas sucessões de variáveis aleatórias, embora um caso particular (variáveis aleatórias definidas por uma sucessão de experiência num esquema de Bernoulli) tivesse sido conhecido anteriormente por De Moivre. (A figura 14.11 representa uma distribuição binomial e a correspondente aproximação normal). Laplace estabeleceu um teorema limite central geral, teorema esse que foi, pela primeira vez, completamente demonstrado pelo matemático russo A. Lyapunov em 1901. Em 1922, J. W. Lindeberg generalizou

o resultado de Laplace provando que a propriedade é satisfeita se as variáveis aleatórias são independentes e admitem uma distribuição comum com a mesma esperança matemática e a mesma variância, seja $E(X_k) = m$ e $\text{Var}(X_k) = \sigma^2$ para todo o k . Neste caso a variável reduzida é

$$T_n = \frac{\sum_{k=1}^n X_k - nm}{\sigma\sqrt{n}}.$$

Lindeberg verificou que a independência por si só não é suficiente para garantir a propriedade limite central, e formulou outra condição (hoje conhecida por *condição de Lindeberg*) a qual, conjuntamente com a independência, é suficiente. Em 1935, W. Feller demonstrou que a condição de Lindeberg é necessária e suficiente para que as variáveis aleatórias independentes satisfaçam à propriedade limite central. Não analisaremos aqui a condição de Lindeberg e apenas nos limitamos a assinalar que ela implica que

$$\text{Var}(S_n) \rightarrow \infty \text{ quando } n \rightarrow \infty.$$

Felizmente, muitas variáveis aleatórias independentes que se definem em problemas práticos verificam automaticamente a condição de Lindeberg e por conseguinte também a propriedade limite central. Até agora a teoria para variáveis aleatórias dependentes está incompleta. Apenas alguns casos particulares têm sido tratados. Muita da investigação contemporânea na teoria das probabilidades está centrada na busca de teoremas gerais dizendo respeito a variáveis aleatórias dependentes.

14.31 Exercícios

1. Demonstrar a desigualdade de Tchebycheff no caso discreto.
2. Se a é qualquer número real, provar que

$$P(|X - a| > c\lambda) \leq \frac{1}{c^2}$$

para todo $c > 0$, com $\lambda^2 = \int_{-\infty}^{\infty} (t - a)^2 f_X(t) dt$. A desigualdade de Tchebycheff corresponde ao caso particular em que $a = E(X)$.

3. Seja X a variável aleatória definida pelo número de ocorrências do acontecimento favorável em n experiências num esquema de Bernoulli; a probabilidade do acontecimento favorável é p . Provar que, para todo $\epsilon > 0$,

$$P\left(\left|\frac{X}{n} - p\right| > \epsilon\right) \leq \frac{1}{4n\epsilon^2}.$$

4. Lança-se ao ar n vezes uma moeda imparcial; o número de ocorrências da face "cara" é X . Determinar o menor n para o qual a desigualdade de Tchebycheff implique

$$P\left(0,4 < \frac{X}{n} < 0,6\right) > 0,90.$$

5. Numa linha de produção sabe-se que o número X de artigos manufacturados defeituosos numa hora tem uma distribuição de Poisson com média $E(X) = 100$. Utilizar a desigualdade de Tchebycheff para calcular um limite inferior para a probabilidade de que uma dada hora se produza um número de artigos defeituosos compreendido entre 90 e 110.
6. Supor que uma variável aleatória X tem uma distribuição normal standard (média 0 e variância 1). Seja p a probabilidade de que X difira de $E(X)$ mais do que três vezes o seu desvio padrão. Aplicar a desigualdade de Tchebycheff para determinar um limite superior para p . Recorrer em seguida a tábuas convenientes da distribuição normal para mostrar que existe um limite superior para p que é aproximadamente $1/50$ da obtida pela desigualdade de Tchebycheff.
7. Dada uma sucessão de variáveis aleatórias independentes X_1, X_2, \dots , cada uma das quais admitindo uma distribuição normal. Seja $m_k = E(X_k)$ e $\sigma_k^2 = \text{Var}(X_k)$. Provar que tal sucessão goza da propriedade limite central. [Sugestão: Recordar o Exercício 7 da Secção 14.24].
8. Sejam X_1, X_2, \dots variáveis aleatórias independentes com a mesma distribuição binomial. Admitimos que cada X_k toma os valores possíveis 0 e 1 com probabilidades $P(X_k = 1) = p$ e $P(X_k = 0) = q$, sendo $p + q = 1$. Seja $Z_n = X_1 + X_2 + \dots + X_n$. A variável aleatória Z_n dá conta do número de ocorrências do acontecimento favorável em n experiências num esquema de Bernoulli.
- (a) Provar que a propriedade limite central toma a forma

$$\lim_{n \rightarrow \infty} P\left(\frac{Z_n - np}{\sqrt{npq}} \leq t\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-u^2/2} du.$$

- (b) Considerar a aproximação sugerida na alínea (a) para estimar a probabilidade de obter entre 45 e 55 vezes a face "cara" se uma moeda imparcial é lançada 100 vezes. Recorrer à tabela 14.1, pg. 603 para o cálculo.
9. Com a notação do Exercício 8, o teorema limite central para variáveis aleatórias definindo uma sucessão de experiências num esquema de Bernoulli pode escrever-se na forma

$$\lim_{n \rightarrow \infty} \frac{P\left(t_1 \leq \frac{Z_n - np}{\sqrt{npq}} \leq t_2\right)}{\Phi(t_2) - \Phi(t_1)} = 1,$$

onde Φ é a distribuição normal standard. Para este caso particular pode provar-se que a fórmula é igualmente válida quando t_1 e t_2 são funções de n dadas por $t_1 = (a - np)/\sqrt{npq}$ e $t_2 = (b - np)/\sqrt{npq}$, com a e b constantes positivas, $a < b$.

- (a) Demonstrar que esta relação implica a fórmula assintótica

$$\sum_{k=a}^b \binom{n}{k} p^k q^{n-k} \sim \Phi\left(\frac{b - np + \frac{1}{2}}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np - \frac{1}{2}}{\sqrt{npq}}\right) \quad \text{com } n \rightarrow \infty.$$

- (b) Lança-se 180 vezes um dado perfeito. Utilizar a aproximação sugerida na alínea (a) para estimar a probabilidade de que sai a face seis precisamente 30 vezes. Para o cálculo recorrer à tabela 14.1, pg. 603.
10. Lança-se 100 vezes um dado perfeito. Utilizar a aproximação sugerida no Exercício 9(a) para estimar a probabilidade de que sai a face seis: (a) precisamente 25 vezes; (b) pelo menos 25 vezes. Recorrer à tabela 14.1, pg. 603 para o cálculo. Recorrer à tabela 14.1, pg. 603.

Bibliografia

1. H. Cramer, *Elements of Probability Theory*, John Wiley, New York, 1955.
2. H. Cramer, *Mathematical Methods of Statistics*, Princeton Univ. Press, Princeton, N.J., 1946.
3. W. Feller, *An Introduction to Probability Theory and Its Applications*, 2.^a edição, John Wiley, N. York, 1957.
4. B. V. Gnedenko e A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables*, Addison-Wesley, Reading, Mars, 1954.
5. S. Goldberg, *Probability, an Introduction*, Prentice-Hall Englewood Cliffs, N. J., 1960.
6. H. Levy e L. Roth, *Elements of Probability*, Oxford Univ. Press, London and New York, 1936.
7. M. Loève, *Probability Theory. Foundations, Random Sequences*. Van Nostrand, New York, 1955.
8. M. E. Munroe, *Theory of Probability*, Mc Graw-Hill, Book Co., Inc., New York, 1951.
9. J. Neyman, *First Course in Probability and Statistics*, Holt Rinehart and Winston, N. York, 1950.
10. E. Parzen, *Modern Probability Theory and Its Applications*, John Wiley, N. York, 1960.
11. I. Todhunter, *A History of the Mathematical Theory of Probability from the Time of Pascal to Laplace*, Chelsea. Publishing Co., N. York, 1949.
12. J. V. Uspensky, *Introduction to Mathematical Probability*, Mc Graw-Hill Book Co., Inc., N. York, 1937.

INTRODUÇÃO À ANÁLISE NUMÉRICA

15.1 Introdução histórica

O planeta Urano foi descoberto em 1781 por um excelente astrónomo amador, William Herschel (1738-1822) com um telescópio de fabrico artesanal. A partir das leis de Kepler foi rapidamente calculada a órbita do planeta Urano, a partir de algumas observações muito separadas entre si. Determinou-se que a distância média do Urano ao Sol era cerca de duas vezes maior que a de Saturno e que o período de uma revolução sideral era de cerca de 84 anos. Por volta de 1830 os dados empíricos acumulados puzeram em evidência desvios inexplicáveis da órbita prevista. Alguns astrónomos chegaram a admitir que a lei da gravitação universal de Newton não seria correcta para distâncias tão grandes como a do planeta Urano ao Sol, outros admitiam que as perturbações seriam devidas a um cometa ainda não descoberto ou a um planeta mais afastado.

Um estudante ainda não bacharelado da Universidade de Cambridge, John Couch Adams (1819-1892), mostrou-se intrigado pela possibilidade de existência de um planeta ainda desconhecido. Decide-se por si próprio à difícil tarefa de calcular a influência de um tal planeta nas posições observadas de Urano, supondo correcta a lei universal da gravitação de Newton. Completou os seus cálculos em 1845 e solicitou ao Observatório Real de Greenwich que fossem feitas pesquisas para detecção de um hipotético planeta, mas os seus pedidos não foram atendidos.

Entretanto independentemente e quase simultaneamente foram efectuados cálculos análogos por Jean Joseph Leverrier (1811-1877) de Paris, que solicitou a Johann Galle, director do Observatório de Berlim, para confirmar as suas previsões. Na mesma noite em que recebeu a carta de Leverrier, Galle descobriu o novo planeta, *Neptuno*, quase na posição calculada. Este era outro triunfo da lei de Newton da gravitação, e um dos primeiros grandes triunfos da *análise numérica*, a arte e a ciência de calcular.

A história da análise numérica data de tempos antigos. Quase cerca de 2.000 anos A.C. os babilónios calcularam tábuas matemáticas. Encontrou-se um pedaço de barro contendo uma tábua dos quadrados dos inteiros de 1 a 60. Os babilónios adoravam os corpos celestes e elaboraram efemérides astronómicas. O famoso astrónomo de Ale-

xandria, Claudio Ptolomeu (cerca de 150 D.C.) possui umas efemérides babilónicas de eclipses que datavam do ano de 747 A.C.

Em 220 A.C., Arquimedes usou os polígonos regulares como aproximações da circunferência e deduziu que $3\frac{1}{7} < \pi < 3\frac{1}{2}$. O trabalho de cálculo numérico desde então até ao século XVII centrou-se principalmente na preparação de tábuas astronómicas. O advento da álgebra no século XVI deu lugar a uma renovada actividade em todos os ramos da Matemática, incluindo a análise numérica. Em 1614, Napier publicou a primeira tábua de lagarismos. Em 1620, os lagarismos do seno e da tangente foram tabulados com sete decimais. Em 1628 estavam calculadas as tábuas de lagarismos decimais dos números 1 a 100.000 com catorze casas decimais.

Os cálculos com séries começaram a desenvolver-se próximo do final do Século XVII, paralelamente ao desenvolvimento do cálculo. Nos começos do Século XVIII Jacob Stirling e Brook Taylor estabeleciam os fundamentos do *Cálculo de diferenças finitas*, o qual desempenha actualmente um papel fundamental na Análise Numérica. Com a previsão da existência e localização do planeta Neptuno por Adams e Leverrier em 1845, a importância científica da análise numérica ficou definida de uma vez para sempre.

Mais tarde no Século XIX o desenvolvimento das máquinas de cálculo automáticas ainda mais estimulou o desenvolvimento da análise numérica. Este desenvolvimento foi explosivo desde o final da II Guerra Mundial devido ao progresso nas máquinas de calcular electrónicas de alta velocidade. Elas tornaram possíveis a obtenção de grande número de resultados científicos que anteriormente pareciam inacessíveis.

A arte de calcular (distinta da ciência de calcular) dá grande importância à elaboração numérica necessária a um cálculo particular. Também trata questões tais como a precisão, os erros, e a verificação dos cálculos. Este aspecto da análise numérica não será discutido aqui; apreende-se melhor ao efectuar os cálculos numéricos em problemas concretos. Para ter um bom conhecimento dos métodos práticos e técnicos a seguir, o leitor deverá consultar os livros sobre análise numérica, alguns dos quais são citados na Bibliografia na parte final do Capítulo. A bibliografia contém também algumas das tábuas matemáticas mais usadas; muitas delas dão também informação prática acerca do modo como proceder num cálculo determinado.

Este capítulo é uma introdução à *ciência* de calcular: Contém alguns dos princípios matemáticos fundamentais que necessita quem trabalhar em análise numérica, quer o faça com um pequeno calculador ou com uma grande máquina de calcular de alta velocidade. À parte o seu valor prático, a matéria deste capítulo tem interesse em si própria e espera-se que esta breve introdução venha a estimular o leitor a aprofundar mais este importante e fascinante ramo da matemática.

15.2 Aproximação polinomial

Uma ideia básica da análise numérica é a de utilizar funções simples, habitualmente polinómios, para aproximar uma dada função f . No Volume I analisámos um tipo de aproximação polinomial ao estudarmos a fórmula de Taylor (Teorema 7.1). O problema consistia na determinação de um polinómio P que coincida com uma dada função f , e algumas das suas derivadas num dado ponto. Provou-se que se f é uma função com

derivada de ordem n num ponto a , existe um e um só polinómio P de grau $\leq n$ o qual verificará as $n+1$ relações

$$P(a) = f(a), \quad P'(a) = f'(a), \quad \dots, \quad P^{(n)}(a) = f^{(n)}(a).$$

A solução é dada pelo *polinómio de Taylor*.

$$P(x) = \sum_{k=0}^n \frac{f^{(k)}(a)}{k!} (x-a)^k.$$

Analisámos também o erro que se comete aproximado $f(x)$ por $P(x)$ em pontos x distintos de a . Este erro define-se pela diferença $E_n(x) = f(x) - P(x)$, pelo que podemos escrever

$$f(x) = \sum_{k=0}^n \frac{f^{(k)}(a)}{k!} (x-a)^k + E_n(x).$$

Para dizer algo mais a respeito do erro necessitamos maior informação a respeito da função f . Por exemplo, se f possui derivada contínua de ordem $n+1$ em algum intervalo contendo a , então para cada x deste intervalo o erro pode exprimir-se através de um integral ou de uma derivada de ordem $n+1$:

$$E_n(x) = \frac{1}{n!} \int_a^x (x-t)^n f^{(n+1)}(t) dt = \frac{f^{(n+1)}(c)}{(n+1)!} (x-a)^{n+1},$$

com c um ponto situado entre a e x . (Ver Secções 7.5 e 7.7 do Volume I.)

Existem várias outras maneiras de aproximar uma dada função f por polinómios, dependendo do uso que deva fazer-se da aproximação. Por exemplo, em vez de pretender determinar um polinómio que coincida com f e com algumas das suas derivadas num ponto dado, podemos pretender determinar um polinómio que tome os mesmos valores que f num certo número de pontos distintos. Especificamente, se os pontos distintos dados são x_0, x_1, \dots, x_n podemos procurar um polinómio P que satisfaça às condições

$$(15.1) \quad P(x_0) = f(x_0), \quad P(x_1) = f(x_1), \quad \dots, \quad P(x_n) = f(x_n).$$

Uma vez que são $n+1$ as condições a verificar tentamos determinar um polinómio de grau $\leq n$, seja

$$P(x) = \sum_{k=0}^n a_k x^k,$$

com $n+1$ coeficientes a_0, a_1, \dots, a_n a serem determinados. As $(n+1)$ condições (15.1) conduzem-nos a um sistema de $(n+1)$ equações lineares cujas incógnitas são

os coeficientes do polinómio. Da teoria das equações lineares pode concluir-se que este sistema tem uma e uma só solução; logo um tal polinómio existe sempre. Se as equações são resolvidas pela regra de Cramer, os coeficientes a_0, a_1, \dots, a_n exprimem-se na forma de cocientes de determinantes. Na prática, todavia, o polinómio P raras vezes se determina deste modo porque os cálculos são extremamente trabalhosos quando n é grande. Métodos mais simples têm sido desenvolvidos para determinar os polinómios de aproximação. Alguns deles serão analisados nas Secções que se seguem. O polinómio que resolve o problema atrás referido chama-se *polinómio interpolador*.

Outro tipo corrente de aproximação polinomial é a *aproximação por mínimos quadrados*. Neste caso a função dada f é definida e integrável num intervalo $[a, b]$ e pretendemos um polinómio P de grau $\leq n$, tal que o erro quadrático médio

$$\int_a^b |f(x) - P(x)|^2 dx$$

seja o menor possível. Na Secção 15.4 provaremos que para uma função contínua f , tal polinómio existe e é único. Os polinómios de Legendre introduzidos na Secção 1.14 desempenham um papel fundamental na resolução deste problema.

15.3. Aproximação polinomial e espaços lineares normados

Todos os diferentes tipos de aproximação polinomial referidos na Secção anterior estão interligados por uma ideia central, a qual é melhor descrita na terminologia própria dos espaços lineares.

Seja V um espaço linear de funções que contém todos os polinómios de grau $\leq n$ e que contém também a função f que se pretende aproximar. Os polinómios constituem um subespaço S de dimensão finita, com $\dim S = n + 1$. Quando falamos de aproximar f por um polinómio P de S , consideramos a diferença $f - P$, que chamamos o erro de aproximação, e tratamos então de determinar um método para medir a grandeza desse erro.

Se V é um espaço euclidiano, então define-se um produto interno (x, y) e a correspondente norma dada por $\|x\| = (x, x)^{1/2}$, e podemos usar a norma $\|f - P\|$ como uma medida da grandeza do erro.

Por vezes podem introduzir-se normas em espaços lineares não euclidianos, isto é, espaços lineares nos quais não se define um produto interno. Estas normas foram introduzidas na Secção 7.26. Por comodidade repetimos aqui a definição.

DEFINIÇÃO DE NORMA. *Seja V um espaço linear. Uma função real N definida em V diz-se uma norma se goza das seguintes propriedades:*

- (a) $N(f) \geq 0$ para todo o f de V .
- (b) $N(cf) = |c|N(f)$ para todo o f em V e todo o escalar c .
- (c) $N(f + g) \leq N(f) + N(g)$ para todo o par de funções f e g de V .
- (d) $N(f) = 0$ implica $f = 0$.

Um espaço linear no qual se define uma norma, diz-se um espaço linear normado.

A norma de f escreve-se algumas vezes $\|f\|$ em vez de $N(f)$. Nesta notação, as propriedades fundamentais acabadas de referir escrevem-se:

- (a) $\|f\| \geq 0$,
- (b) $\|cf\| = |c|\|f\|$,
- (c) $\|f + g\| \leq \|f\| + \|g\|$,
- (d) $\|f\| = 0$ implica $f = 0$.

Uma função N que satisfaz a (a), (b), (c), mas não a (d), diz-se uma semi-norma. Alguns problemas da teoria da aproximação tratam com espaços lineares semi-normados; outros com espaços lineares normados. Os exemplos seguintes serão discutidos neste capítulo.

EXEMPLO 1. Semi-norma de Taylor. Para um inteiro fixo $n \geq 1$, seja V o espaço linear das funções que admitem derivada num ponto fixo a até à ordem n . Se $f \in V$, seja

$$N(f) = \sum_{k=0}^n |f^{(k)}(a)|.$$

É fácil verificar que a função N assim definida é uma semi-norma. Não é uma norma porque $N(f) = 0$ se e só se

$$f(a) = f'(a) = \dots = f^{(n)}(a) = 0,$$

e estas equações podem satisfazer-se com uma função não nula. Por exemplo, $N(f) = 0$ quando $f(x) = (x - a)^{n+1}$.

EXEMPLO 2. Semi-norma de interpolação. Seja V o espaço linear de todas as funções reais definidas num intervalo $[a, b]$. Para um conjunto fixo de $n + 1$ pontos distintos x_0, x_1, \dots, x_n em $[a, b]$, seja N definida pela equação

$$N(f) = \sum_{k=0}^n |f(x_k)|$$

se $f \in V$. Esta função N é uma semi-norma em V . Não é uma norma porque $N(f) = 0$ se e só se $f(x_0) = f(x_1) = \dots = f(x_n) = 0$, e é evidente que essas equações podem verificar-se com uma função f não nula em $[a, b]$.

EXEMPLO 3. Norma quadrática. Seja C o espaço linear das funções contínuas num intervalo $[a, b]$. Se $f \in C$ definimos

$$(15.2) \quad N(f) = \left(\int_a^b |f(x)|^2 dx \right)^{1/2}.$$

Esta é uma norma consequência do produto interno

$$(f, g) = \int_a^b f(x) \overline{g(x)} dx.$$

Nota: Seja S o conjunto das funções f integráveis em $[a, b]$. O conjunto S é um espaço linear, e a função N definida por (15.2) é uma semi-norma de S . Não é uma norma porque se tem $N(f) = 0$ sem que f seja identicamente nula em $[a, b]$.

EXEMPLO 4. Norma maximal. Seja C o espaço linear das funções contínuas num intervalo $[a, b]$. Se $f \in C$, define-se

$$N(f) = \max_{a \leq x \leq b} |f(x)|,$$

onde o símbolo do segundo membro representa o máximo absoluto de $|f|$ em $[a, b]$. A verificação das quatro propriedades da norma é pedida no Exercício 4 da Secção 15.5.

15.4 Problemas fundamentais da aproximação polinomial.

Seja C o espaço das funções contínuas num dado intervalo $[a, b]$, e S o subespaço linear de todos os polinómios de grau $\leq n$. Suponhamos também que se define uma norma ou semi-norma em C . Escolhamos uma função f em C . Se existir um polinómio P em S tal que

$$\|f - P\| \leq \|f - Q\|$$

para todos os polinómios Q em S , dizemos que P é a *melhor aproximação polinomial* para f com o grau especificado. A expressão “o melhor” é naturalmente, relativa à norma dada (ou semi-norma). O melhor polinómio para uma escolha da norma não o será necessariamente para outra norma escolhida.

Uma vez escolhida a norma ou a semi-norma, três problemas se apresentam de imediato.

1. *Existência.* Dada a função f em C , existirá um melhor polinómio de aproximação para f com grau previamente determinado?

2. *Unicidade.* Se existir esse polinómio f será ele único?

3. *Construção.* Se existir esse polinómio óptimo para f com o grau previamente dado, como poderá determinar-se?

Existem, sem dúvida, muitos outros problemas que podem ser considerados. Por exemplo, se o único melhor polinómio de aproximação P_n de grau $\leq n$ existir, podemos pretender obter limites superiores para $\|f - P_n\|$, os quais poderão ser utilizados para satisfazer a requisitos práticos. Ou podemos averiguar se $\|f - P_n\| \rightarrow 0$ quando $n \rightarrow \infty$ para a norma dada ou provavelmente para outra norma. Se for assim, dizemos que os polinómios de aproximação convergem para f nesta norma. Em tal hipótese existem aproximações tão próximas quanto se queira relativas a essa norma se n é suficientemente grande. Estes exemplos ilustram alguns dos tipos de problemas que se estudam na teoria geral da aproximação polinomial. Neste estudo preliminar dedicamos a nossa

atenção especialmente aos três problemas de existência, unicidade e construção, como já nos referimos.

No caso da aproximação mediante polinómios de Taylor estes problemas podem resolver-se totalmente. Se f admite derivadas até à ordem n num ponto a , é fácil provar que o polinómio da melhor aproximação de grau $\leq n$ relativo à semi-norma de Taylor para este n é o polinómio de Taylor

$$(15.3) \quad P(x) = \sum_{k=0}^n \frac{f^{(k)}(a)}{k!} (x-a)^k.$$

Com efeito, para este polinómio temos

$$\|f - P\| = \sum_{k=0}^n |f^{(k)}(a) - P^{(k)}(a)| = 0,$$

pelo que a desigualdade $\|f - P\| \leq \|f - Q\|$ é trivialmente satisfeita para todo o polinómio Q . Deste modo P é o polinómio da melhor aproximação relativamente a esta semi-norma. Para provar a unicidade, consideremos qualquer polinómio Q de grau $\leq n$ tal que $\|f - Q\| = 0$. Esta equação implica que

$$Q(a) = f(a), \quad Q'(a) = f'(a), \quad \dots, \quad Q^{(n)}(a) = f^{(n)}(a).$$

Pelo teorema 7.1 do Volume I sabemos que o polinómio de Taylor em (15.3) é o único polinómio satisfazendo a estas equações. Portanto $Q = P$. A equação (15.3) resolve também o problema da construção.

Todos os três problemas podem ser resolvidos para qualquer norma deduzida de um produto interno. Neste caso, o teorema 1.16 diz-nos que existe um só polinómio em S para o qual a norma $\|f - P\|$ é a mínima possível. Com efeito, este P é a projecção de f em S e vem dado por uma fórmula explícita,

$$P(x) = \sum_{k=0}^n (P, e_k) e_k(x),$$

onde e_0, e_1, \dots, e_n são funções que constituem uma base ortonormada para S .

Por exemplo, se for C o espaço das funções reais contínuas no intervalo $[-1, 1]$ e se

$$(f, g) = \int_{-1}^1 f(x)g(x) dx,$$

os polinómios de Legendre normalizados $\varphi_0, \varphi_1, \dots, \varphi_n$ formam uma base ortonormada para S , e a projecção f_n de f sobre S vem dada por

$$f_n(x) = \sum_{k=0}^n (f, \varphi_k) \varphi_k(x), \quad \text{onde} \quad (f, \varphi_k) = \int_{-1}^1 f(t) \varphi_k(t) dt.$$

Recordemos que os polinómios de Legendre normalizados são dados por

$$\varphi_k(x) = \sqrt{\frac{2k+1}{2}} P_k(x), \quad \text{com} \quad P_k(x) = \frac{1}{2^k k!} \frac{d^k}{dx^k} (x^2 - 1)^k.$$

Os seis primeiros polinómios normalizados são

$$\begin{aligned} \varphi_0(x) &= \sqrt{\frac{1}{2}}, & \varphi_1(x) &= \sqrt{\frac{3}{2}} x, & \varphi_2(x) &= \frac{1}{2} \sqrt{\frac{5}{2}} (3x^2 - 1), & \varphi_3(x) &= \frac{1}{2} \sqrt{\frac{7}{2}} (5x^3 - 3x), \\ \varphi_4(x) &= \frac{1}{8} \sqrt{\frac{9}{2}} (35x^4 - 30x^2 + 3), & \varphi_5(x) &= \frac{1}{8} \sqrt{\frac{11}{2}} (63x^5 - 70x^3 + 15x). \end{aligned}$$

Os problemas correspondentes à semi-norma serão tratados na Secção 15.6. Nas últimas secções analizaremos a aproximação polinomial relativa à norma maximal.

15.5 Exercícios

- Provar que cada uma das seguintes famílias de funções é um espaço linear.
 - Todos os polinómios.
 - Todos os polinómios de grau $\leq n$.
 - Todas as funções contínuas num intervalo I .
 - Todas as funções deriváveis num intervalo I .
 - Todas as funções deriváveis até à ordem n num intervalo I .
 - Todas as funções admitindo derivadas até à ordem n num ponto fixo x_0 .
 - Todas as funções admitindo desenvolvimentos em série de potências na vizinhança de um dado ponto x_0 .
- Dizer se sim ou não cada uma das seguintes famílias de funções reais é um espaço linear.
 - Todos os polinómios de grau n .
 - Todas as funções definidas e limitadas num intervalo $[a, b]$.
 - Todas as funções em escada definidas num intervalo $[a, b]$.
 - Todas as funções monótonas num intervalo $[a, b]$.
 - Todas as funções integráveis num intervalo $[a, b]$.
 - Todas as funções seccionalmente monótonas num intervalo $[a, b]$.
 - Todas as funções que podem exprimir-se na forma $f - g$, com f e g monótonas crescentes num intervalo $[a, b]$.
- Seja C o espaço linear das funções reais contínuas num intervalo $[a, b]$. Uma função N define-se em C na forma que a seguir se indica. Determinar para cada caso quais das quatro propriedades da norma que são satisfeitas, e determinar em consequência disso se N é uma norma, uma semi-norma, ou nenhuma delas.

(a) $N(f) = f(a)$.	(e) $N(f) = \left \int_a^b f(x) dx \right $.
(b) $N(f) = f(a) $.	(f) $N(f) = \int_a^b f(x) dx$.
(c) $N(f) = f(b) - f(a) $.	(g) $N(f) = \int_a^b f(x) ^2 dx$.
(d) $N(f) = \int_a^b f(x) dx$.	(h) $N(f) = \left \int_a^b f(x) dx \right ^2$.

4. Seja C o espaço linear normado das funções contínuas num intervalo $[a, b]$. Se $f \in C$, definir

$$N(f) = \max_{a \leq x \leq b} |f(x)|.$$

Provar que N é uma norma para C .

5. Seja B o espaço linear de todas as funções reais definidas e limitadas num intervalo $[a, b]$. Se $f \in B$, definir

$$N(f) = \sup_{a \leq x \leq b} |f(x)|,$$

onde o símbolo do segundo membro representa o supremo (menor limite superior) do conjunto de todos os números $|f(x)|$ para x em $[a, b]$. Provar que N é uma norma para B . Esta chama-se a *norma do supremo*.

6. Considerar o Exercício 3. Determinar quais das funções N dadas têm a propriedade de $N(fg) \leq N(f)N(g)$ para todo o par de funções f e g em C .
7. Para um inteiro fixo $n \geq 1$, seja S o conjunto de todas as funções admitindo derivada de ordem n num ponto fixo x_0 . Se $f \in S$, seja

$$N(f) = \sum_{k=0}^n \frac{1}{k!} |f^{(k)}(x_0)|.$$

- (a) Provar que N é uma semi-norma em S .
- (b) Provar que $N(fg) \leq N(f)N(g)$ para todo o par f, g de S .
Provar também que a semi-norma de Taylor não goza desta propriedade.
8. Seja f uma função real contínua no intervalo $[-1, 1]$.
- (a) Provar que o melhor polinómio quadrático de aproximação relativo à norma quadrática em $[-1, 1]$ é dado por

$$P(x) = \frac{1}{2} \int_{-1}^1 f(t) dt + \frac{3}{2} x \int_{-1}^1 t f(t) dt + \frac{5}{8} (3x^2 - 1) \int_{-1}^1 (3t^2 - 1) f(t) dt.$$

- (b) Determinar uma fórmula análoga para o melhor polinómio de aproximação de grau ≤ 4 .
9. Calcular as constantes a, b, c de tal modo que o integral $\int_{-1}^1 |e^x - (a + bx + cx^2)|^2 dx$ seja tão pequeno quanto possível.
10. Seja $f(x) = |x|$ para $-1 \leq x \leq 1$. Determinar o polinómio de grau ≤ 4 que melhor aproxima f em $[-1, 1]$ relativamente à norma quadrática.
11. Seja C o espaço linear das funções contínuas reais em $[a, b]$ com o produto interno $(f, g) = \int_a^b f(x)g(x)dx$. Seja e_0, \dots, e_n uma base ortonormada para o subespaço S dos polinómios de grau $\leq n$. Seja P o polinómio em S que melhor aproxima uma dada função f em C relativamente à norma quadrática.
- (a) Provar que o quadrado da norma do erro é dado por

$$\|f - P\|^2 = \|f\|^2 - \sum_{k=0}^n (f, e_k)^2.$$

- (b) Calcular este erro explicitamente quando $[a, b] = [-1, 1]$, $n = 2$, e $f(x) = |x|$.
12. Seja $f(x) = 1/x$ para $x \neq 0$.
- (a) Provar que o polinómio constante P que melhor aproxima f num intervalo $[1, n]$, em relação à norma quadrática é $P(x) = (\log n)/(n-1)$. Calcular $\|P - f\|^2$ para este P .
- (b) Determinar o polinómio linear de aproximação P que melhor aproxima f num intervalo $[1, n]$ em relação à norma quadrática. Calcular $\|P - f\|^2$ para este P quando $n = 2$.

13. Seja $f(x) = e^x$.

(a) Provar que o polinômio constante P que melhor aproxima f no intervalo $[0, n]$ relativamente à norma quadrática é $P(x) = (e^n - 1)/n$. Calcular $\|P - f\|^2$ para este P .

(b) Determinar o polinômio linear P que melhor aproxima f no intervalo $[0, 1]$ em relação à norma quadrática. Calcular $\|P - f\|^2$ para este P .

14. Sejam P_0, P_1, \dots, P_n , $n + 1$ polinômios ortogonais em $[a, b]$ relativamente ao produto interno do Exercício 11. Supor também igualmente que, P_k tem grau k .

(a) Provar que três polinômios consecutivos quaisquer desse conjunto estão ligados pela fórmula de recorrência

$$P_{k+1}(x) = (a_k x + b_k)P_k(x) + c_k P_{k-1}(x)$$

para $1 \leq k \leq n-1$, onde a_k, b_k, c_k são constantes.

(b) Determinar esta relação por recorrência explicitamente quando os polinômios são os polinômios de Legendre ortonormados.

15. Considerar o Exercício 14 e representar por p_k os coeficientes de x^k em $P_k(x)$.

(a) Provar que $a_k = p_{k+1}/p_k$.

(b) Recorrer à fórmula de recorrência do Exercício 14 para estabelecer a fórmula

$$\sum_{k=0}^m P_k(x)P_k(y) = \frac{p_m}{p_{m+1}} \frac{P_{m+1}(x)P_m(y) - P_m(x)P_{m+1}(y)}{x - y},$$

válida para $x \neq y$. Analizar o caso limite quando $x = y$.

15.6. Polinômios interpoladores

Voltamos agora à aproximação mediante polinômios. Conhecem-se os valores de uma função f em $n + 1$ pontos distintos x_0, x_1, \dots, x_n e pretendemos determinar um polinômio P de grau $\leq n$ que verifica as condições

$$(15.4) \quad P(x_0) = f(x_0), \quad P(x_1) = f(x_1), \quad \dots, \quad P(x_n) = f(x_n).$$

Em primeiro lugar provamos que se tal polinômio existir é único. Em seguida provaremos que esse polinômio existe por meio da sua construção explícita. Este polinômio minimiza a distância de f a P , medida pela semi-norma de interpolação para este n ,

$$\|f - P\| = \sum_{k=0}^n |f(x_k) - P(x_k)|.$$

Visto que a distância é 0 se P verifica (15.4), o polinômio interpolador P é a melhor aproximação relativa a esta semi-norma.

TEOREMA 15.1. TEOREMA DE UNICIDADE. *Se dados $n + 1$ pontos distintos x_0, x_1, \dots, x_n , P e Q são dois polinômios de grau $\leq n$ tais que*

$$P(x_k) = Q(x_k)$$

para cada $k = 0, 1, 2, \dots, n$, então $P(x) = Q(x)$ para todo x .

Demonstração. Seja $R(x) = P(x) - Q(x)$. A função R é um polinómio de grau $\leq n$ que tem $n + 1$ zeros distintos nos pontos x_0, x_1, \dots, x_n . O único polinómio que goza desta propriedade é o polinómio nulo. Deste modo $R(x) = 0$ para todo x , pelo que $P(x) = Q(x)$ para todo x .

O polinómio interpolador P pode construir-se de diferentes modos. Vamos expor em primeiro lugar o método de Lagrange. Seja $A(x)$ o polinómio dado por

$$(15.5) \quad A(x) = (x - x_0)(x - x_1) \cdots (x - x_n) = \prod_{j=0}^n (x - x_j).$$

Este polinómio tem um zero simples em cada um dos pontos x_j . Seja $A_k(x)$ o polinómio de grau n obtido de $A(x)$ por supressão de factor $x - x_k$; isto é,

$$(15.6) \quad A_k(x) = \prod_{\substack{j=0 \\ j \neq k}}^n (x - x_j).$$

O polinómio $A_k(x)$ tem um zero simples em cada ponto $x_j \neq x_k$. No próprio ponto x_k tem-se

$$(15.7) \quad A_k(x_k) = \prod_{\substack{j=0 \\ j \neq k}}^n (x_k - x_j).$$

Este é não nulo porque nenhum factor do produto é nulo. Portanto o polinómio $A_k(x)/A_k(x_k)$ tem o valor 1 quando $x = x_k$ e o valor 0 quando $x = x_j$, para $x_j \neq x_k$. Seja agora

$$P(x) = \sum_{k=0}^n \frac{f(x_k)A_k(x)}{A_k(x_k)}.$$

Quando $x = x_j$, cada parcela desta soma anula-se excepto o termo de ordem j , que tem o valor $f(x_j)$. Por conseguinte $P(x_j) = f(x_j)$ para cada j . Visto que cada parcela desta soma é um polinómio de grau n , a própria soma é um polinómio de grau $\leq n$. Assim, encontrámos um polinómio satisfazendo às condições requeridas. Este resultado pode resumir-se no teorema seguinte:

TEOREMA 15.2. *Dados $n + 1$ pontos distintos x_0, x_1, \dots, x_n e $n + 1$ números reais $f(x_0), f(x_1), \dots, f(x_n)$, não necessariamente distintos, existe um e um só polinómio P de grau $\leq n$ tal que $P(x_j) = f(x_j)$ para $j = 0, 1, 2, \dots, n$. Este polinómio é dado pela fórmula*

$$(15.8) \quad P(x) = \sum_{k=0}^n \frac{f(x_k)A_k(x)}{A_k(x_k)},$$

onde $A_k(x)$ é o polinómio definido por (15.6).

A fórmula (15.8) que dá $P(x)$ chama-se a *fórmula de interpolação de Lagrange*. Podemos escrevê-la do seguinte modo

$$P(x) = \sum_{k=0}^n f(x_k) L_k(x),$$

onde $L_k(x)$ é um polinómio de grau n dado por

$$(15.9) \quad L_k(x) = \frac{A_k(x)}{A_k(x_k)}.$$

Assim, para cada x fixo, $P(x)$ é uma combinação linear dos valores fixados $f(x_0), f(x_1), \dots, f(x_n)$. Os multiplicadores $L_k(x)$ dependem unicamente dos pontos x_0, x_1, \dots, x_n e não dos valores fixados. Chamam-se os *coeficientes de interpolação de Lagrange*. Se utilizarmos as fórmulas (15.6) e (15.7) podemos escrever (15.9) na forma

$$(15.10) \quad L_k(x) = \prod_{\substack{j=0 \\ j \neq k}}^n \frac{x - x_j}{x_k - x_j}.$$

Esta fórmula proporciona um método eficiente para calcular o número $L_k(x)$ para um dado x .

Nota: Os coeficientes de Lagrange $L_k(x)$ exprimem-se muitas vezes na forma

$$L_k(x) = \frac{A_k(x)}{A'(x_k)},$$

onde A' é a derivada do polinómio (15.5). Para demonstrar esta fórmula basta provar que $A'(x_k) = A_k(x_k)$. Derivando a igualdade $A(x) = (x - x_k) A_k(x)$ obtemos $A'(x) = (x - x_k) A'_k(x) + A_k(x)$. Quando $x = x_k$ isto dá-nos $A'(x_k) = A_k(x_k)$.

EXEMPLO. Determinar o polinómio de grau ≤ 3 que toma os valores y_0, y_1, y_2, y_3 nos pontos $-2, -1, 1, 2$ respectivamente.

Resolução. Tomamos $x_0 = -2, x_1 = -1, x_2 = 1, x_3 = 2$. Os polinómios $L_k(x)$ em (15.10) são dados por

$$L_0(x) = \frac{(x+1)(x-1)(x-2)}{(-2+1)(-2-1)(-2-2)} = -\frac{1}{12}(x+1)(x-1)(x-2),$$

$$L_1(x) = \frac{(x+2)(x-1)(x-2)}{(-1+2)(-1-1)(-1-2)} = \frac{1}{6}(x+2)(x-1)(x-2),$$

$$L_2(x) = \frac{(x+2)(x+1)(x-2)}{(1+2)(1+1)(1-2)} = -\frac{1}{6}(x+2)(x+1)(x-2),$$

$$L_3(x) = \frac{(x+2)(x+1)(x-1)}{(2+2)(2+1)(2-1)} = \frac{1}{12}(x+2)(x+1)(x-1).$$

Portanto o polinómio pedido é

$$\begin{aligned} P(x) &= y_0 L_0(x) + y_1 L_1(x) + y_2 L_2(x) + y_3 L_3(x) \\ &= -\frac{y_0}{12}(x+1)(x-1)(x-2) + \frac{y_1}{6}(x+2)(x-1)(x-2) \\ &\quad - \frac{y_2}{6}(x+2)(x+1)(x-2) + \frac{y_3}{12}(x+2)(x+1)(x-1). \end{aligned}$$

Para calcular o valor de $P(x)$ para um determinado x é usualmente preferível deixar o polinómio nesta forma em vez de o ordenar segundo as potências crescentes de x . Por exemplo, se $y_0 = -5$, $y_1 = 1$, $y_2 = 1$, e $y_3 = 7$, o valor de $P(x)$ para $x = \frac{1}{2}$ vem dado por

$$\begin{aligned} P\left(\frac{1}{2}\right) &= \frac{5}{12}\left(\frac{5}{2}\right)\left(\frac{1}{2}\right)\left(-\frac{1}{2}\right) + \frac{1}{6}\left(\frac{7}{2}\right)\left(\frac{1}{2}\right)\left(-\frac{1}{2}\right) - \frac{1}{6}\left(\frac{3}{2}\right)\left(\frac{5}{2}\right)\left(-\frac{1}{2}\right) + \frac{7}{12}\left(\frac{7}{2}\right)\left(\frac{5}{2}\right)\left(\frac{1}{2}\right) \\ &= -\frac{25}{96} - \frac{7}{48} + \frac{5}{48} + \frac{245}{96} = \frac{278}{96} = \frac{27}{8}. \end{aligned}$$

15.7 Pontos de interpolação igualmente separados

Na proposição que acabámos de efectuar os pontos de interpolação x_0, x_1, \dots, x_n supõem-se distintos, mas fora disso são arbitrários. Suponhamos agora que estão igualmente espaçados e mostremos que os coeficientes de Lagrange $L_k(x)$ se simplificam consideravelmente. Seja $x_0 < x_1 < x_2 < \dots < x_n$ e represente h distância entre pontos consecutivos. Então podemos escrever

$$x_j = x_0 + jh$$

para $j = 0, 1, 2, \dots, n$. Visto que $x_k - x_j = (k - j)h$, (15.10) escreve-se

$$(15.11) \quad L_k(x) = \prod_{\substack{j=0 \\ j \neq k}}^n \frac{x - x_0 - jh}{(k - j)h} = \prod_{\substack{j=0 \\ j \neq k}}^n \frac{t - j}{k - j},$$

onde

$$t = \frac{x - x_0}{h}.$$

No último termo do segundo membro de (15.11) o produto de factores independentes de t é

$$\begin{aligned} (15.12) \quad \prod_{\substack{j=0 \\ j \neq k}}^n \frac{1}{k - j} &= \left(\prod_{j=0}^{k-1} \frac{1}{k - j} \right) \left(\prod_{j=k+1}^n \frac{1}{k - j} \right) = \frac{1}{k!} \prod_{j=k+1}^n \frac{(-1)}{j - k} \\ &= \frac{(-1)^{n-k}}{k! (n - k)!} = \frac{(-1)^{n-k}}{n!} \binom{n}{k}, \end{aligned}$$

onde $\binom{n}{k}$ é o coeficiente binomial. Visto que $x = x_0 + th$, (5.11) escreve-se agora

$$(15.13) \quad L_k(x_0 + th) = \frac{(-1)^{n-k}}{n!} \binom{n}{k} \prod_{\substack{j=0 \\ j \neq k}}^n (t - j).$$

Para cada n fixo, o segundo membro de (15.13) é uma função de k e t que pode ser tabulada. O National Bureau of Standards preparou tábuas de coeficientes de Lagrange para pontos de interpolação igualmente espaçados. (Ver referência 13 da Bibliografia, no final do presente capítulo.) Se x e h se escolhem de maneira que o número $t = (x - x_0)/h$ seja tal que para ele estejam tabulados os coeficientes de Lagrange $L_k(x_0 + th)$, o cálculo de $P(x_0 + th)$ reduz-se a uma multiplicação de $f(x_k)$ pelo valor tabulado $L_k(x_0 + th)$, seguida de adição.

15.8. Análise do erro na interpolação polinomial

Sejam f uma função definida num intervalo $[a, b]$ contendo os $n + 1$ pontos distintos x_0, x_1, \dots, x_n e P o polinômio interpolador de grau $\leq n$ que coincide com f nestes pontos. Se alterarmos os valores de f em pontos diferentes dos pontos de interpolação não se alterará o polinômio P . Isto mostra que a função f e o polinômio P podem diferir consideravelmente em pontos distintos dos pontos de interpolação. Se a função dada f tem certas qualidades de "regularidade" em todo o intervalo $[a, b]$ podemos esperar que o polinômio interpolador P será uma boa aproximação de f em pontos diferentes dos x_k . O teorema seguinte dá-nos uma expressão útil que nos permite estudar o erro na interpolação polinomial quando a função dada admite derivada de ordem $n + 1$ em $[a, b]$.

TEOREMA 15.3. *Sejam x_0, x_1, \dots, x_n , $n + 1$ pontos distintos no domínio da função f e P o polinômio interpolador de grau $\leq n$ que coincide com f naqueles pontos. Escolha-se um ponto x no domínio de f e seja $[\alpha, \beta]$ qualquer intervalo fechado contendo os pontos x_0, x_1, \dots, x_n e x . Se f possui uma derivada de ordem $n + 1$ no intervalo $[\alpha, \beta]$ existe pelo menos um ponto c no intervalo aberto (α, β) tal que*

$$(15.14) \quad f(x) - P(x) = \frac{A(x)}{(n + 1)!} f^{(n+1)}(c),$$

com

$$A(x) = (x - x_0)(x - x_1) \cdots (x - x_n).$$

Nota: O ponto c depende de x e de n .

Demonstração. Se for x um dos pontos de interpolação x_k , então $A(x_k) = 0$ e a equação (15.14) é trivialmente satisfeita para qualquer c em (α, β) . Suponhamos então, que x não é um dos pontos de interpolação. Consideremos x fixo e definamos uma nova função F em $[\alpha, \beta]$ mediante a equação

$$(15.15) \quad F(t) = A(x)[f(t) - P(t)] - A(t)[f(x) - P(x)].$$

O segundo membro desta igualdade, com função de t , admite derivada de ordem $n+1$, logo o mesmo é verdadeiro para o primeiro membro. Porque $P(t)$ é um polinómio em t de grau $\leq n$, a sua derivada de ordem $n+1$ é identicamente nula. O polinómio $A(t)$ tem grau $n+1$, sendo o termo de ordem mais elevada t^{n+1} , e tem-se $A^{(n+1)}(t) = (n+1)!$. Por conseguinte, se derivamos ambos os membros de (15.15) $n+1$ vezes em relação a t obtemos a fórmula

$$(15.16) \quad F^{(n+1)}(t) = A(x)f^{(n+1)}(t) - (n+1)! [f(x) - P(x)].$$

Da definição de F em (15.15) vê-se que F se anula nos $n+1$ pontos de interpolação x_0, x_1, \dots, x_n e também no ponto x . Deste modo $F(t) = 0$ em $n+2$ pontos distintos no intervalo $[\alpha, \beta]$. Estes pontos determinam $n+1$ subintervalos de $[\alpha, \beta]$ adjacentes, anulando-se a função F nos pontos extremos de cada um desses subintervalos. Pelo teorema de Rolle, a derivada $F'(t)$ deve anular-se em, pelo menos, um ponto t anterior a cada subintervalo. Se escolhermos precisamente um tal t de cada subintervalo obtemos $n+1$ pontos distintos no intervalo aberto (α, β) nos quais $F'(t) = 0$. Estes pontos, por sua vez determinam n subintervalos em cujas extremidades se tem $F'(t) = 0$. Aplicando o teorema de Rolle a F' concluímos que a segunda derivada $F''(t)$ se anula em pelo menos n pontos distintos de (α, β) . Depois da aplicação, deste modo, do teorema de Rolle $n+1$ vezes, encontramos finalmente que existe pelo menos um ponto c em (α, β) no qual $F^{(n+1)}(c) = 0$. Substituindo este valor de c na equação (15.16) obtemos

$$(n+1)! [f(x) - P(x)] = A(x)f^{(n+1)}(c),$$

a qual coincide com (15.14). Está pois demonstrado o teorema.

Deve observar-se que, tal como na aproximação com polinómios de Taylor, o erro contém a derivada de ordem $n+1$, $f^{(n+1)}(c)$, calculada num ponto desconhecido c . Se os valores extremos de $f^{(n+1)}$ em $[\alpha, \beta]$ são conhecidos, podem obter-se limites superior e inferior para o erro.

Suponhamos agora que os pontos de interpolação estão igualmente espaçados e que $x_0 < x_1 < \dots < x_n$. Se h representa a distância entre cada par de pontos de interpolação consecutivos, podemos escrever

$$x_j = x_0 + jh \quad \text{e} \quad x = x_0 + th,$$

onde $t = (x - x_0)/h$. Uma vez que $x - x_j = (t - j)h$, o polinómio $A(x)$ pode escrever-se

$$A(x) = \prod_{j=0}^n (x - x_j) = h^{n+1} \prod_{j=0}^n (t - j).$$

A fórmula (15.14) vem

$$(15.17) \quad f(x) - P(x) = \frac{f^{(n+1)}(c)}{(n+1)!} h^{n+1} \prod_{j=0}^n (t - j),$$

com $t = (x - x_0)/h$.

EXEMPLO. *O erro na interpolação linear.* Suponhamos que uma função f , com segunda derivada, está tabulada e que pretendemos calcular o seu valor num ponto x compreendido entre duas entradas consecutivas x_0 e $x_0 + h$. Se usarmos a interpolação linear aproximamos o gráfico de f relativo ao intervalo $[x_0, x_0 + h]$ por uma recta, como se mostra na figura 15.1. Se P representa o polinómio linear interpolador, o erro estimado em (15.17) vem

$$(15.18) \quad f(x) - P(x) = \frac{f''(c)}{2!} h^2 t(t-1),$$

onde $t = (x - x_0)/h$. Quando x está entre x_0 e $x_0 + h$ tem-se $0 < t < 1$ e o valor máximo de $|t(t-1)|$ neste intervalo é $\frac{1}{4}$. Deste modo (15.18) dá-nos a estimativa

$$|f(x) - P(x)| \leq \frac{|f''(c)| h^2}{8}.$$

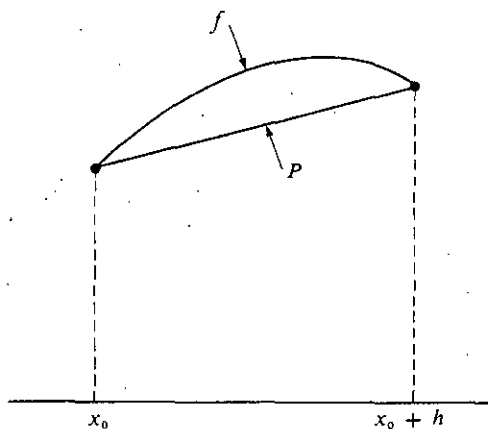


FIG. 15.1. Interpolação linear.

O ponto c é desconhecido no intervalo $(x_0, x_0 + h)$. Se a derivada segunda de f for limitada neste intervalo, seja $|f''(x)| \leq M$, a estimativa do erro vem

$$|f(x) - P(x)| \leq \frac{Mh^2}{8}.$$

Em particular, se f for um seno ou um cosseno, então $|f''(x)| \leq 1$ para todo x e tem-se $|f(x) - P(x)| \leq h^2/8$. Numa tabela de senos e cossenos com entradas espaçadas de 1 grau (um grau = $\pi/180$ radianos) tem-se $h = \pi/180$, pelo que

$$\frac{h^2}{8} = \frac{\pi^2}{8(180)^2} < \frac{10}{259,200} < \frac{1}{25,000} = 0,00004.$$

porque este erro não excede $\frac{1}{2}$ da quarta casa decimal, a interpolação linear poderia considerar-se como dando um resultado satisfatório numa tábua com quatro decimais. O erro calculado pode melhorar-se nos pontos da tábua em que $|f'''(c)|$ seja consideravelmente menor que 1.

5.9. Exercícios

1. Para cada alínea determinar o polinómio P de menor grau possível satisfazendo às condições dadas.

- (a) $P(-1) = 0$, $P(0) = 2$, $P(2) = 7$.
- (b) $P(1) = 1$, $P(2) = 0$, $P(3) = 0$, $P(4) = 1$.
- (c) $P(1) = 1$, $P(2) = 2$, $P(3) = 3$, $P(0) = 1$.
- (d) $P(0) = -2$, $P(1) = 0$, $P(-1) = -2$, $P(2) = 16$.
- (e) $P(-2) = 11$, $P(-1) = -11$, $P(0) = -5$, $P(1) = -1$.

2. Seja $f(x) = \cos(\pi x/4)$. Determinar o polinómio de menor grau que tome os mesmos valores que f nos pontos $-2, -\frac{1}{2}, 0, \frac{1}{2}, 2$.

3. Seja P um polinómio de grau $\leq n$ e $A(x) = (x - x_0)(x - x_1) \cdots (x - x_n)$, sendo x_0, x_1, \dots, x_n , $n+1$ pontos distintos.

- (a) Provar que, qualquer que seja o polinómio B , o polinómio Q definido por $Q(x) = P(x) + A(x)B(x)$ coincide com P nos pontos x_0, x_1, \dots, x_n .
- (b) Provar também o recíproco, isto é, se Q é um polinómio qualquer que coincide com P nos pontos x_0, x_1, \dots, x_n , então $Q(x) = P(x) + A(x)B(x)$ para algum polinómio B .

4. (a) Determinar o polinómio Q de menor grau possível que verifique as condições

$$Q(-2) = -5, \quad Q(-1) = -1, \quad Q(1) = 1, \quad Q'(0) = -1.$$

[Sugestão: Determinar em primeiro lugar um polinómio P que toma os valores indicados em $-2, -1, 1$ e utilizar em segundo o Exercício 3 para determinar Q].

(b) Determinar o polinómio Q de menor grau possível que verifique as condições da alínea (a) com $Q'(0) = -3$ em vez de $Q'(0) = -1$.

5. Seja $f(x) = \log_2 x$ para $x > 0$. Calcular $P(32)$, com P o polinómio de menor grau que coincide com f nos pontos

- (a) $x = 1, 64$.
- (b) $x = 1, 16, 256$.
- (c) $x = 4, 16, 64$.
- (d) $x = 1, 4, 16, 64, 256$.

Para cada caso calcular a diferença $f(32) - P(32)$. Estes exemplos mostram que a precisão na interpolação polinomial não é necessariamente melhorada pelo aumento do número de pontos de interpolação.

6. Os coeficientes de interpolação de Lagrange $L_k(x)$ dados pela equação (15.10) dependem não somente de x , mas também dos pontos de interpolação x_0, x_1, \dots, x_n . Podemos expressar esta dependência escrevendo $L_k(x) = L_k(x; N)$, representando N o vector no espaço com $n+1$ dimensões definido por $N = (x_0, x_1, \dots, x_n)$. Para um dado número real b , seja b o vector no espaço com $n+1$ dimensões com todas as componentes iguais a b . Se $a \neq 0$, provar que

$$L_k(ax + b; aX + b) = L_k(x; X).$$

Esta é a chamada *propriedade de invariância* dos coeficientes de interpolação de Lagrange. O exercício que se segue mostra que pode utilizar-se esta propriedade para auxiliar na simplificação dos cálculos na prática.

7. Seja P um polinómio de grau ≤ 4 que toma os valores

$$P(2.4) = 72, \quad P(2.5) = 30, \quad P(2.7) = 18, \quad P(2.8) = 24, \quad P(3.0) = 180.$$

(a) Introduzir novos pontos de interpolação u_j relacionados com os pontos x dados pela fórmula $u_j = 10x_j - 24$. Os u_j são inteiros. Para cada $k = 0, 1, 2, 3, 4$, determinar os coeficientes de interpolação de Lagrange $L_k(x)$ em função de u , com $u = 10x - 24$.

(b) Utilizar a propriedade da invariância do Exercício 6 para calcular $P(2, 6)$.

8. Uma tábua da função $f(x) = \log x$ contém entradas desde $x = 1$ até $x = 10$ com intervalos de 0,001. Os valores da função para valores de x intermédios de cada par de entradas consecutivas calculam-se por interpolação linear. Supor que as entradas na tábua são exactas.

(a) Provar que o erro na interpolação linear não excederá $\frac{1}{2}$ na sexta casa decimal.

(b) Para que os valores de x será a interpolação linear satisfatória para uma tábua de sete decimais?

(c) Qual deveria ser o espaçamento das entradas no intervalo $1 \leq x \leq 2$ de maneira que a interpolação linear fosse satisfatória numa tábua de sete decimais?

Nos Exercícios 9 a 15, x_0, x_1, \dots, x_n são pontos distintos e

$$A(x) = \prod_{j=0}^n (x - x_j), \quad A_k(x) = \prod_{\substack{j=0 \\ j \neq k}}^n (x - x_j), \quad L_k(x) = \frac{A_k(x)}{A_k(x_k)}.$$

9. Deduzir a fórmula $A'(x) = \sum_{k=0}^n A_k'(x)$ recorrendo (a) à derivação do logaritmo; (b) à fórmula de interpolação de Lagrange.

10. Demonstrar cada uma das fórmulas seguintes:

$$(a) \sum_{k=0}^n L_k(x) = 1 \quad e \quad \sum_{k=0}^n \frac{A'_k(x)}{A'_k(x_k)} = 0 \quad \text{para todo } x.$$

$$(b) \sum_{k=0}^n \frac{1}{A'_k(x_k)} = 0. \quad \text{[Sugestão: Utilizar a alínea (a) com valores adequados de } x]$$

11. Seja P qualquer polinómio de grau $\leq n$. Mostrar que o coeficiente de x^n é igual a

$$\sum_{k=0}^n \frac{P(x_k)}{A'_k(x_k)}.$$

12. (a) Determinar a e b de modo que o polinómio

$$P_k(x) = \{a + b(x - x_k)\} L_k(x)^2$$

tenha as seguintes propriedades:

$$P_k(x_i) = 0 \text{ para todo } i, \quad P'_k(x_k) = 1, \quad e \quad P'_k(x_i) = 0 \text{ para } i \neq k.$$

(b) Determinar c e d de modo que o polinómio

$$Q_k(x) = \{c + d(x - x_k)\}L_k(x)^2$$

tenha as seguintes propriedades:

$$Q_k(x_k) = 1, \quad Q_k(x_i) = 0 \quad \text{para } i \neq k, \quad \text{e} \quad Q'_k(x_i) = 0 \quad \text{para } i \neq k.$$

(c) Seja $H(x) = \sum_{k=0}^n f(x_k)Q_k(x) + \sum_{k=0}^n f'(x_k)P_k(x)$, onde f é uma função dada derivável em x_0, x_1, \dots, x_n . Provar que

$$H(x_i) = f(x_i) \quad \text{e} \quad H'(x_i) = f'(x_i) \quad \text{para } i.$$

Demonstrar também que existe quando muito um polinómio $H(x)$ de grau $\leq 2n + 1$ com esta propriedade.

13. (a) Sejam P e Q dois polinómios de grau $\leq n$ satisfazendo às $n + 1$ condições

$$P(x_0) = Q(x_0), \quad P'(x_1) = Q'(x_1), \quad P''(x_2) = Q''(x_2), \quad \dots, \quad P^{(n)}(x_n) = Q^{(n)}(x_n).$$

Provar que $P(x) = Q(x)$ para todo x .

(b) Seja $B_0(x) = 1$, e para $n \geq 1$ defina-se

$$B_n(x) = \frac{x(x-n)^{n-1}}{n!}.$$

Provar que $B'_n(x) = B_{n-1}(x-1)$ para $n \geq 1$ e deduzir que

$$B_n(0) = B'_n(1) = B''_n(2) = \dots = B^{(n-1)}_n(n-1) = 0 \quad \text{e} \quad B^{(n)}_n(n) = 1.$$

(c) Mostrar que o único polinómio de grau $\leq n$ que satisfaz às condições

$$P(0) = c_0, \quad P'(1) = c_1, \quad P''(2) = c_2, \quad \dots, \quad P^{(n)}(n) = c_n$$

vem dado por

$$P(x) = \sum_{k=0}^n c_k B_k(x).$$

(d) Se $x_k = x_0 + kh$ para $k = 0, 1, 2, \dots, n$, onde $h > 0$, generalizar os resultados de (b) e (c).

14. Supondo que x_0, x_1, \dots, x_n são inteiros que satisfazem a $x_0 < x_1 < \dots < x_n$.

(a) Provar que $|A'(x_k)| \geq k!(n-k)!$ e demonstrar que

$$\sum_{k=0}^n \frac{1}{|A'(x_k)|} \leq \frac{2^n}{n!}.$$

(b) Seja P um qualquer polinómio de grau n , com o termo de maior grau igual a x^n . Represente M o maior dos números $|P(x_0)|, |P(x_1)|, \dots, |P(x_n)|$. Provar que $M \geq n!/2^n$. [Sugestão: Usar a alínea (a) do Exercício 11].

15. Demonstrar as fórmulas seguintes. Nas alíneas (a) e (b), x é qualquer ponto diferente de x_0, x_1, \dots, x_n .

$$(a) \frac{A'(x)}{A(x)} = \sum_{j=0}^n \frac{1}{x - x_j}.$$

$$(b) \frac{A''(x)}{A'(x)} = \frac{A_k(x)}{A'(x)} \sum_{\substack{j=0 \\ j \neq k}}^n \frac{1}{x - x_j} + \sum_{\substack{j=0 \\ j \neq k}}^n \frac{1}{x - x_j} - \frac{A(x)}{A'(x)} \sum_{\substack{j=0 \\ j \neq k}}^n \frac{1}{(x - x_j)^2}.$$

$$(c) \frac{A''(x_k)}{A'(x_k)} = 2 \sum_{\substack{j=0 \\ j \neq k}}^n \frac{1}{x_k - x_j}.$$

16. Seja $P_n(x)$ o polinómio de grau $\leq n$ que coincide com a função $f(x) = e^{ax}$ para os valores inteiros de $x = 0, 1, \dots, n$. Visto que este polinómio depende de a represente-se por $P_n(x; a)$. Provar que o limite

$$\lim_{a \rightarrow 0} \frac{P_n(x; a) - 1}{a}$$

existe e é um polinómio em x . Determinar este polinómio explicitamente.

15.10. Fórmula de interpolação de Newton

Seja P_n o polinómio interpolador de grau $\leq n$ que coincide com uma dada função f em $n+1$ pontos distintos x_0, x_1, \dots, x_n . A fórmula de interpolação de Lagrange diz-nos que

$$P_n(x) = \sum_{k=0}^n L_k(x) f(x_k),$$

onde $L_k(x)$ é um polinómio de grau n (o coeficiente de interpolação de Lagrange) dado pelo produto

$$(15.19) \quad L_k(x) = \prod_{\substack{j=0 \\ j \neq k}}^n \frac{x - x_j}{x_k - x_j}, \quad \text{para } k = 0, 1, 2, \dots, n.$$

Suponhamos que juntamos um novo ponto de interpolação x_{n+1} aos pontos dados x_0, x_1, \dots, x_n . Para determinar o correspondente polinómio P_{n+1} pela fórmula de Lagrange é necessário calcular um novo coeficiente de interpolação L_{n+1} e calcular de novo todos os anteriores coeficientes L_0, L_1, \dots, L_n , sendo agora cada um deles um polinómio de grau $n+1$. Na prática isto implica um certo volume de trabalho. Consequentemente é desejável ter outra fórmula para determinar P_n , a qual permita uma passagem mais fácil de P_n para P_{n+1} . Uma tal fórmula foi descoberta por Newton; deduzi-la-emos a partir do seguinte teorema.

TEOREMA 15.4. *Dados $n+2$ pontos distintos $x_0, x_1, \dots, x_n, x_{n+1}$, seja P_n o polinómio de grau $\leq n$ que coincide com uma dada função f em x_0, x_1, \dots, x_n e P_{n+1} o polinómio de grau $\leq n+1$ que coincide com f em $x_0, x_1, \dots, x_n, x_{n+1}$. Existe então uma constante c_{n+1} , univocamente determinada por f e pelos pontos de interpolação x_0, \dots, x_{n+1} tal que*

$$(15.20) \quad P_{n+1}(x) = P_n(x) + c_{n+1}(x - x_0) \cdots (x - x_n).$$

Demonstração. Seja $Q(x) = P_n(x) + c(x - x_0) \cdots (x - x_n)$, com c uma constante não especificada. Então Q é um polinómio de grau $\leq n+1$ que coincide com P_n e portanto com f nos $n+1$ pontos x_0, \dots, x_n . Escolhamos agora c de maneira que Q tenha o valor de f também em x_{n+1} . Tal exigirá que

$$f(x_{n+1}) = P_n(x_{n+1}) + c(x_{n+1} - x_0) \cdots (x_{n+1} - x_n).$$

Uma vez que o coeficiente de c é diferente de zero, esta equação tem uma única solução que designamos por c_{n+1} . Fazendo $c = c_{n+1}$ vemos que $Q = P_{n+1}$.

O teorema seguinte exprime $P_n(x)$ em função de c_1, \dots, c_n .

TEOREMA 15.5. FÓRMULA DE INTERPOLAÇÃO DE NEWTON. *Se x_0, \dots, x_n são distintos, tem-se*

$$(15.21) \quad P_n(x) = f(x_0) + \sum_{k=1}^n c_k(x - x_0) \cdots (x - x_{k-1}).$$

Demonstração. Definimos $P_0(x) = f(x)$ e tomamos $n = 0$ em (15.20) para obtermos

$$P_1(x) = f(x_0) + c_1(x - x_0).$$

Façamos agora $n = 1$ em (15.20) para obtermos

$$P_2(x) = P_1(x) + c_2(x - x_0)(x - x_1) = f(x_0) + c_1(x - x_0) + c_2(x - x_0)(x - x_1).$$

Por indução, obtemos (15.21).

A propriedade da fórmula de Newton expressa por (15.20) permite-nos calcular P_{n+1} apenas pela adição de um novo termo a P_n . Esta propriedade não a possui a fórmula de Lagrange.

A utilidade da fórmula de Newton depende, sem dúvida, da facilidade com que os coeficientes c_1, c_2, \dots, c_n , podem ser calculados. O teorema seguinte mostra que c_n é uma combinação linear dos valores $f(x_0), \dots, f(x_n)$.

TEOREMA 15.6. *Os coeficientes da fórmula de interpolação de Newton são dados por*

$$(15.22) \quad c_n = \sum_{k=0}^n \frac{f(x_k)}{A_k(x_k)}, \quad \text{com} \quad A_k(x_k) = \prod_{\substack{j=0 \\ j \neq k}}^n (x_k - x_j).$$

Demonstração. Pela fórmula de Lagrange tem-se

$$P_n(x) = \sum_{k=0}^n L_k(x) f(x_k),$$

sendo $L_k(x)$ o polinómio de grau n dado por (15.19). Visto que o coeficiente de x^n em $L_k(x)$ é $1/A_k(x_k)$, o coeficiente de x^n em $P_n(x)$ é a soma definida em (15.22). Por outro

lado, a fórmula de Newton mostra que o coeficiente de x^n em $P_n(x)$ é igual a c_n , o que completa a demonstração.

A equação (15.22) proporciona um método directo de cálculo dos coeficientes na fórmula de Newton. Os números $A_k(x_k)$ aparecem também como factores no denominador dos coeficientes de interpolação de Lagrange $L_k(x)$. Na próxima Secção analisamos um outro método de cálculo dos coeficientes quando a função é tabulada para intervalos iguais.

15.11. Pontos de interpolação igualmente espaçados. O operador das diferenças sucessivas

No caso de pontos de interpolação igualmente espaçados com $x_k = x_0 + kh$ para $k = 0, 1, \dots, n$, podemos utilizar a equação (15.12) para obtermos

$$\frac{1}{A_k(x_k)} = \prod_{\substack{j=0 \\ j \neq k}}^n \frac{1}{x_k - x_j} = \frac{1}{h^n} \prod_{\substack{j=0 \\ j \neq k}}^n \frac{1}{k - j} = \frac{(-1)^{n-k}}{n! h^n} \binom{n}{k}.$$

Neste caso a fórmula que define c_n no teorema 15.6 vem

$$(15.23) \quad c_n = \frac{1}{n! h^n} \sum_{k=0}^n (-1)^{n-k} \binom{n}{k} f(x_k),$$

A soma do segundo membro pode calcular-se doutro modo por intermédio de um operador Δ chamado o *operador das diferenças sucessivas*.

DEFINIÇÃO. *Seja h um número real fixo e f uma função dada. A função Δf definida pela equação*

$$\Delta f(x) = f(x + h) - f(x)$$

chama-se a primeira diferença de f . Define-se naqueles pontos x para os quais ambos x e $x + h$ pertencem ao domínio de f . As diferenças de ordem superior $\Delta^2 f, \Delta^3 f, \dots$ definem-se por recorrência do modo seguinte:

$$\Delta^{k+1}f = \Delta(\Delta^k f) \quad \text{para } k = 1, 2, 3, \dots$$

Nota: As notações $\Delta_h f(x)$ e $\Delta f(x; h)$ são também usadas para $\Delta f(x)$ em especial quando se pretende expressar a dependência de h . É conveniente definir $\Delta^0 f = f$.

A diferença de ordem n , $\Delta^n f(x)$, é uma combinação linear dos valores $f(x), f(x + h), \dots, f(x + nh)$ da função. Por exemplo, tem-se

$$\begin{aligned} \Delta^2 f(x) &= \{f(x + 2h) - f(x + h)\} - \{f(x + h) - f(x)\} \\ &= f(x + 2h) - 2f(x + h) + f(x). \end{aligned}$$

Em geral tem-se

$$(15.24) \quad \Delta^n f(x) = \sum_{k=0}^n (-1)^{n-k} \binom{n}{k} f(x + kh).$$

Isto pode ser facilmente demonstrado por indução em n , recorrendo à lei do triângulo de Pascal para os coeficientes binomiais:

$$\binom{n}{k-1} + \binom{n}{k} = \binom{n+1}{k}.$$

Suponhamos agora f definida em $n+1$ pontos igualmente espaçados $x_k = x_0 + kh$ para $k = 0, 1, 2, \dots, n$. Então de (15.23) e (15.24) obtemos a fórmula

$$c_n = \frac{1}{h^n n!} \Delta^n f(x_0).$$

Isto proporciona um método rápido para o cálculo dos coeficientes na fórmula de interpolação de Newton. O diagrama da tábuia 15.1, chamada uma *tábuia de diferenças*, mostra como as diferenças sucessivas podem ser sistematicamente calculadas a partir de valores tabulados de f em pontos igualmente espaçados. Na tábuia escreve-se f_k em vez de $f(x_k)$.

A fórmula de interpolação de Newton (15.21) escreve-se agora

$$(15.25) \quad P_n(x) = f(x_0) + \sum_{k=1}^n \frac{\Delta^k f(x_0)}{k! h^k} \prod_{j=0}^{k-1} (x - x_j).$$

TÁBUA 15.1

x	$f(x)$	$\Delta f(x)$	$\Delta^2 f(x)$	$\Delta^3 f(x)$
x_0	f_0			
	f_1	$f_1 - f_0 = \Delta f(x_0)$		
x_1	f_1		$\Delta f(x_1) - \Delta f(x_0) = \Delta^2 f(x_0)$	
	f_2	$f_2 - f_1 = \Delta f(x_1)$		$\Delta^2 f(x_1) - \Delta^2 f(x_0) = \Delta^3 f(x_0)$
x_2	f_2		$\Delta f(x_2) - \Delta f(x_1) = \Delta^2 f(x_1)$	
	f_3	$f_3 - f_2 = \Delta f(x_2)$		
x_3	f_3			

Se escrevermos

$$\prod_{j=0}^{k-1} (x - x_j) = \prod_{j=0}^{k-1} (x - x_0 - jh) = h^k \prod_{j=0}^{k-1} \left(\frac{x - x_0}{h} - j \right) = h^k \prod_{j=0}^{k-1} (t - j),$$

onde $t = (x - x_0)/h$, a fórmula (15.25) vem

$$(15.26) \quad P_n(x) = f(x_0) + \sum_{k=1}^n \frac{\Delta^k f(x_0)}{k!} \prod_{j=0}^{k-1} (t - j).$$

15.12. Polinômios factoriais

O produto $t(t-1)\cdots(t-k+1)$ que aparece na soma em (15.26) é um polinômio em t de grau k chamado o *polinômio factorial*, ou a *potência factorial de ordem k de t* . Representa-se pelo símbolo $t^{(k)}$. Então, por definição,

$$t^{(k)} = \prod_{j=0}^{k-1} (t - j).$$

Definimos também $t^{(0)} = 1$. Se considerarmos o operador diferença Δ com $h = 1$, isto é $\Delta f(x) = f(x+1) - f(x)$, encontramos que

$$\Delta t^{(n)} = n t^{(n-1)} \quad \text{para } n \geq 1.$$

Esta é uma fórmula análoga à fórmula de derivação de potências, $Dt^n = n t^{n-1}$. Assim, a potência factorial $t^{(n)}$ está relacionada com as diferenças do mesmo modo que as potências ordinárias t^n o estão com as derivadas.

Com o uso dos polinômios factoriais, a fórmula de interpolação de Newton (15.26) escreve-se

$$P_n(x_0 + th) = \sum_{k=0}^n \frac{\Delta^k f(x_0)}{k!} t^{(k)}.$$

Posta nesta forma, a fórmula de Newton assemelha-se à fórmula de Taylor para o polinômio de grau $\leq n$ que coincide com f e as suas n primeiras derivadas em x_0 . Se escrevermos

$$\binom{t}{k} = \frac{t^{(k)}}{k!} = \frac{t(t-1)\cdots(t-k+1)}{k!},$$

a fórmula de Newton toma a forma

$$P_n(x_0 + th) = \sum_{k=0}^n \binom{t}{k} \Delta^k f(x_0).$$

Outras propriedades dos polinômios factoriais serão desenvolvidas nos exercícios que se seguem.

15.13. Exercícios

1. Seja $\Delta f(x) = f(x+h) - f(x)$. Se f é um polinômio de grau n , por exemplo

$$f(x) = \sum_{r=0}^n a_r x^r$$

com $a_n \neq 0$, demonstrar que: (a) $\Delta^k f(x)$ é um polinômio de grau $n-k$ se $k \leq n$; (b) $\Delta^n f(x) = n! h^n a_n$; (c) $\Delta^k f(x) = 0$ para $k > n$.

2. Seja $\Delta f(x) = f(x+h) - f(x)$. Se $f(x) = \sin(ax+b)$, provar que

$$\Delta^n f(x) = \left(2 \sin \frac{ah}{2}\right)^n \sin \left(ax + b + \frac{nah + n\pi}{2}\right).$$

3. Seja $\Delta f(x) = f(x+h) - f(x)$.

(a) Se $f(x) = a^x$, com $a > 0$, provar que $\Delta^k f(x) = (a^h - 1)^k a^x$.

(b) Se $g(x) = (1+a)^{x/h}$, com $a > 0$, provar que $\Delta^k g(x) = a^k g(x)$.

(c) Mostrar que o polinômio P_n de grau n que toma os valores $P_n(k) = (1+a)^k$ para $k = 0, 1, 2, \dots, n$ é dado por

$$P_n(x) = \sum_{k=0}^n \frac{a^k}{k!} x^{(k)}.$$

4. Seja $x^{(n)}$ a potência factorial de ordem n de x . Posto que $x^{(n)}$ é um polinômio em x de grau n com o valor 0 quando $x = 0$, podemos escrever

$$x^{(n)} = \sum_{k=1}^n S_{k,n} x^k.$$

Os números $S_{k,n}$ chamam-se *números de Stirling de primeira espécie*. Da definição de $x^{(n)}$ é evidente que $S_{n,n} = 1$ para $n \geq 0$.

(a) Provar que $S_{n-1,n} = -n(n-1)/2$ e que $S_{1,n} = (-1)^{n-1}(n-1)!$ para $n \geq 1$.

(b) Provar que $S_{k,n+1} = S_{k-1,n} - nS_{k,n}$. Usar esta relação para verificar as entradas na Tábua 15.2, uma tábua de números de Stirling de primeira espécie, e construir as três colunas seguintes da tábua.

TÁBUA 15.2

n	$S_{1,n}$	$S_{2,n}$	$S_{3,n}$	$S_{4,n}$	$S_{5,n}$	$S_{6,n}$	$S_{7,n}$
1	1						
2	-1	1					
3	2	-3	1				
4	-6	11	-6	1			
5	24	-50	35	-10	1		
6	-120	274	-225	85	-15	1	
7	720	-1764	1624	-735	175	-21	1

(c) Exprimir o polinômio $x^{(4)} + 3x^{(3)} + 2x^{(2)} + 1$ como uma combinação linear de potências de x .

5. (a) Provar que

$$x = x^{(1)}, \quad x^2 = x^{(1)} + x^{(2)}, \quad x^3 = x^{(1)} + 3x^{(2)} + x^{(3)},$$

e que, em geral,

$$x^n = \sum_{k=1}^n \frac{\Delta^k f(0)}{k!} x^{(k)},$$

com $f(x) = x^n$ e $\Delta f(x) = f(x+1) - f(x)$. Os números $T_{k,n} = \Delta^k f(0)/k!$ chamam-se *números de Stirling de segunda espécie*.

(b) Provar

$$\Delta^k x^{n+1} = (x+k) \Delta^k x^n + k \Delta^{k-1} x^n$$

e, a partir dela, provar que $T_{k,n+1} = T_{k-1,n} + k T_{k,n}$.

(c) Verificar a tabela 15.3 dos números de Stirling de segunda espécie pela fórmula de recorrência da alínea (b) e construir as três colunas seguintes da tabela.

TÁBUA 15.3.

n	$T_{1,n}$	$T_{2,n}$	$T_{3,n}$	$T_{4,n}$	$T_{5,n}$	$T_{6,n}$	$T_{7,n}$
1	1						
2	1	1					
3	1	3	1				
4	1	7	6	1			
5	1	15	25	10	1		
6	1	31	90	65	15	1	
7	1	63	301	350	140	21	1

(d) Expressar o polinômio $x^4 + 3x^3 + 2x - 1$ como uma combinação linear de polinômios fatoriais:

6. (a) Se p é um inteiro positivo e se a e b são inteiros com $a < b$, provar que

$$\sum_{k=a}^{b-1} k^{(p)} = \frac{b^{(p+1)} - a^{(p+1)}}{p+1}.$$

Esta fórmula é análoga à da integração $\int_a^b x^p dx$. Deve observar-se, todavia, que o limite superior na soma é $b-1$ e não b .

(b) Verificar que $k(k+3) = 4k^{(1)} + k^{(2)}$. Utilizar a alínea (a) para provar que

$$\sum_{k=1}^n k(k+3) = 4 \frac{(n+1)^{(2)}}{2} + \frac{(n+1)^{(3)}}{3} = \frac{n(n+1)(n+5)}{3}.$$

(c) Se $f(k)$ é um polinómio em k de grau r , provar que

$$\sum_{k=1}^n f(k)$$

é um polinómio em n de grau $r + 1$.

7. Utilizar o método sugerido no Exercício 6 para exprimir cada uma das seguintes somas como um polinómio em n .

$$(a) \sum_{k=1}^n (4k^2 + 7k + 6).$$

$$(c) \sum_{k=1}^n k(k+1)(k+2).$$

$$(b) \sum_{k=1}^n k^2(k+1).$$

$$(d) \sum_{k=1}^n k^4.$$

8. Represente A o operador linear definido por

$$A(f) = a_0 \Delta^n f + a_1 \Delta^{n-1} f + \cdots + a_{n-1} \Delta f + a_n f,$$

com a_0, a_1, \dots, a_n constantes. Este é o chamado *operador diferença com coeficientes constantes*. É análogo ao operador derivação com coeficientes constantes, referido na Secção 6.7. A cada tal A podemos associar o polinómio característica p_A definido por

$$p_A(r) = a_0 r^n + a_1 r^{n-1} + \cdots + a_{n-1} r + a_n.$$

Reciprocamente, a todo o polinómio p podemos associar um operador A do qual é polinómio característico. Se A e B são operadores diferença com coeficientes constantes e se λ é um número real, definir $A + B$, AB e λA mediante fórmulas análogas às usadas na Secção 6.7 para os operadores derivação. Provar então que o teorema 6.6 é válido para operadores diferença com coeficientes constantes.

15.14. Um problema de número relativo à norma maximal

Consideremos um problema que resulta naturalmente da teoria da interpolação polinomial. No teorema 15.3 deduzimos uma expressão para o erro

$$(15.27) \quad f(x) - P(x) = \frac{A(x)}{(n+1)!} f^{(n+1)}(c),$$

onde

$$A(x) = (x - x_0)(x - x_1) \cdots (x - x_n).$$

Aqui P é o único polinómio de grau $\leq n$ que coincide com f em $n + 1$ pontos distintos x_0, x_1, \dots, x_n de $[a, b]$. A função f supõe-se ter derivada até à ordem $n + 1$ em $[a, b]$, e c é um ponto desconhecido situado em $[a, b]$. Para estimar o erro em (15.27) necessitamos conhecer limites para a derivada de ordem $n + 1$, $f^{(n+1)}$, e para o produto $A(x)$. Uma vez que A é um polinómio, o seu valor absoluto tem um máximo em algum ponto do intervalo $[a, b]$. Este máximo dependerá da escolha dos pontos x_0, x_1, \dots, x_n , é natural procurar escolher esses pontos de maneira que a norma A seja a menor possível.

Podemos designar este máximo por $\|A\|$, com $\|A\|$ a norma maximal, definida por

$$\|A\| = \max_{a \leq x \leq b} |A(x)|.$$

O problema consiste em determinar um polinómio do grau especificado que minimiza $\|A\|$. Este problema foi resolvido pela primeira vez por Tchebycheff; a sua solução conduz a uma interessante classe de polinómios que também se apresentam noutras questões. Em primeiro lugar vamos dar algumas indicações relativas e logo de seguida voltaremos ao problema de mínimo referido.

15.15. Polinómios de Tchebycheff

Seja $x + iy$ um número complexo de valor absoluto 1. Pelo desenvolvimento do binómio de Newton temos

$$(x + iy)^n = \sum_{k=0}^n \binom{n}{k} x^{n-k} (iy)^k$$

com n inteiro ≥ 0 . Nesta fórmula escrevemos $x = \cos \theta$, $y = \sin \theta$ e consideramos a parte real de cada membro. Porque

$$(x + iy)^n = (\cos \theta + i \sin \theta)^n = e^{in\theta} = \cos n\theta + i \sin n\theta,$$

a parte real do primeiro membro é $\cos n\theta$. A parte real do segundo membro é a soma referente a todos os valores de k . Daqui resulta

$$(15.28) \quad \cos n\theta = x^n - \binom{n}{2} x^{n-2} y^2 + \binom{n}{4} x^{n-4} y^4 - + \dots$$

Porque $y^2 = \sin^2 \theta = 1 - \cos^2 \theta = 1 - x^2$, o segundo membro de (15.28) é um polinómio em x de grau n . Este polinómio chama-se o polinómio de Tchebycheff de primeira espécie e representa-se por $T_n(x)$.

DEFINIÇÃO. O polinómio de Tchebycheff, $T_n(x)$ define-se, para todo o valor real de x , pela fórmula

$$T_n(x) = \sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n}{2k} x^{n-2k} (x^2 - 1)^k.$$

De (15.28) resulta o seguinte

TEOREMA 15.7. Se $-1 \leq x \leq 1$ tem-se

$$T_n(x) = (n \operatorname{arc} \cos x).$$

Demonstração. Se $\theta = \arccos x$ então $x = \cos \theta$ e $T_n(x) = \cos n\theta$.

Os polinômios de Tchebycheff podem efectivamente calcular-se tomando a parte real de $(x + iy)^n$ com $y^2 = 1 - x^2$, ou então considerando a seguinte fórmula de recorrência.

TEOREMA 15.8. *Os polinômios de Tchebycheff satisfazem à seguinte fórmula de recorrência.*

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x) \quad \text{para } n \geq 1,$$

com $T_0(x) = 1$ e $T_1(x) = x$.

Demonstração. Em primeiro lugar supomos $-1 \leq x \leq 1$ e fazemos $x = \cos \theta$ na identidade trigonométrica

$$\cos(n+1)\theta + \cos(n-1)\theta = 2 \cos \theta \cos n\theta.$$

Isto prova que $T_{n+1}(x) + T_{n-1}(x) = 2xT_n(x)$ para x no intervalo $-1 \leq x \leq 1$. Mas porque ambos os membros são polinômios, esta relação deve verificar-se para qualquer x .

Os cinco polinômios seguintes são

$$\begin{aligned} T_2(x) &= 2x^2 - 1, & T_3(x) &= 4x^3 - 3x, & T_4(x) &= 8x^4 - 8x^2 + 1, \\ T_5(x) &= 16x^5 - 20x^3 + 5x, & T_6(x) &= 32x^6 - 48x^4 + 18x^2 - 1. \end{aligned}$$

A fórmula de recorrência mostra que todos os coeficientes de $T_n(x)$ são inteiros; além disso, o coeficiente de x^n é 2^{n-1} .

O teorema seguinte prova que $T_n(x)$ tem precisamente n zeros de primeira ordem e que estão todos no intervalo $[-1, 1]$.

TEOREMA 15.9. *Se $n \geq 1$ o polinômio $T_n(x)$ tem zeros nos n pontos*

$$x_k = \cos \frac{(2k+1)\pi}{2n}, \quad k = 0, 1, 2, \dots, n-1.$$

Por conseguinte $T_n(x)$ admite a factorização

$$T_n(x) = 2^{n-1}(x - x_0)(x - x_1) \cdots (x - x_{n-1}) = 2^{n-1} \prod_{k=0}^{n-1} \left(x - \cos \frac{(2k+1)\pi}{2n} \right).$$

Demonstração. Usamos a fórmula $T_n(x) = \cos n\theta$. Porque $\cos n\theta = 0$ somente se $n\theta$ é um múltiplo ímpar de $\pi/2$, temos $T_n(x) = 0$ para x em $[-1, 1]$ somente se $n \arccos x = (2k+1)\pi/2$ para algum inteiro k . Deste modo os zeros de T_n no intervalo $[-1, 1]$ encontram-se entre os números

$$(15.29) \quad x_k = \cos \frac{2k+1}{n} \frac{\pi}{2}, \quad k = 0, \pm 1, \pm 2, \dots$$

Os valores $k = 0, 1, 2, \dots, n-1$ dão n zeros distintos x_0, x_1, \dots, x_{n-1} , situados todos no intervalo aberto $(-1, 1)$. Visto que um polinômio de grau n não pode ter mais do que n zeros, esses devem ser *todos* os zeros de T_n . Os restantes x_k em (15.29) são repetições destes n .

TEOREMA 15.10. *No intervalo $[-1, 1]$ os valores extremos de $T_n(x)$ são $+1$ e -1 , alcançados alternadamente nos $n+1$ pontos*

$$(15.30) \quad t_k = \cos \frac{k\pi}{n}, \quad \text{para } k = 0, 1, 2, \dots, n.$$

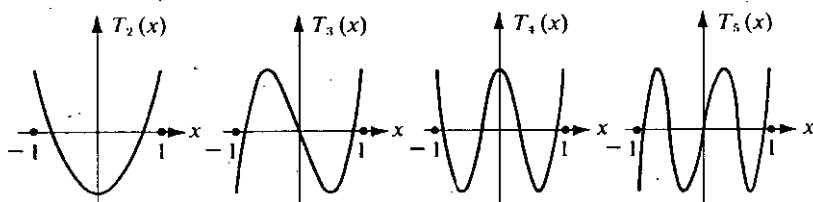


FIG. 15.2. Gráficos dos polinômios de Tchebycheff no intervalo $[-1, 1]$.

Demonstração. Pelo teorema de Rolle, os máximos e mínimos relativos de T_n ocorrerão entre zeros consecutivos; existem $n-1$ tais pontos no intervalo aberto $(-1, 1)$. Da fórmula do cosseno para T_n vê-se que os valores extremos, ± 1 , são atingidos nos $n-1$ pontos interiores $\cos(k\pi/n)$, $k = 1, 2, \dots, n-1$, e também nos extremos $x = 1$ e $x = -1$. Por conseguinte no intervalo fechado $[-1, 1]$ os valores extremos $+1$ e -1 são alcançados *alternadamente* nos $n+1$ pontos t_0, t_1, \dots, t_n dados por $t_k = \cos(k\pi/n)$ para $k = 0, 1, 2, \dots, n$.

A figura 15.2 representa os gráficos de T_2, \dots, T_5 no intervalo $[-1, 1]$.

15.16. Uma propriedade de mínimo dos polinômios de Tchebycheff

Voltamos agora ao problema da determinação de um polinômio com um grau previamente fixado, para o qual a norma maximal é o menor possível. O problema é resolvido pelo seguinte teorema.

TEOREMA 15.11. *Seja $p_n(x) = x^n + \dots$ qualquer polinômio de grau $n \geq 1$ com o coeficiente do termo de menor grau igual a 1, e seja*

$$\|p_n\| = \max_{-1 \leq x \leq 1} |p_n(x)|.$$

Tem-se então a desigualdade

$$(15.31) \quad \|p_n\| \geq \|\tilde{T}_n\|,$$

onde $\tilde{T}_n(x) = T_n(x)/2^{n-1}$. Além disso, a igualdade em (15.31) verifica-se se $p_n = \tilde{T}_n$.

Demonstração. No intervalo $[-1, 1]$ o polinómio \tilde{T}_n toma os seus valores extremos $1/2^{n-1}$ e $-1/2^{n-1}$, alternadamente nos $n+1$ pontos distintos t_k de (15.30). Consequentemente,

Demonstremos em seguida que a desigualdade

$$(15.32) \quad \|p_n\| < \frac{1}{2^{n-1}}$$

nos conduz a uma contradição. Suponhamos, então, que p_n satisfaz a (15.32) e consideremos a diferença

$$r(x) = \tilde{T}_n(x) - p_n(x).$$

Nos pontos t_k dados por (15.30) tem-se

$$r(t_k) = \frac{(-1)^k}{2^{n-1}} - p_n(t_k) = (-1)^k \left[\frac{1}{2^{n-1}} - (-1)^k p_n(t_k) \right].$$

Devido a (15.32) a expressão do parêntesis é positiva. Por conseguinte, $r(t_k)$ tem sinais alternados nos $n+1$ pontos t_0, t_1, \dots, t_n . Uma vez que r é contínua deve anular-se pelo menos uma vez entre duas mudanças de sinal consecutivas. Portanto r tem pelo menos n zeros distintos. Mas porque r é um polinómio de grau $\leq n-1$, isto significa que r é identicamente nulo. Deste modo $P_n = \tilde{T}_n$, logo $\|P_n\| = \|\tilde{T}_n\| = 1/2^{n-1}$, contradizendo (15.32). Isto prova que devemos ter $\|p_n\| \geq 1/2^{n-1} = \|\tilde{T}_n\|$.

Embora o teorema 15.11 se refira ao intervalo $[-1, 1]$ e a um polinómio com o "primeiro" coeficiente 1, pode usar-se para deduzir um resultado correspondente para um intervalo arbitrário $[a, b]$ e um polinómio arbitrário.

TEOREMA 15.12. *Seja $q_n(x) = c_n x^n + \dots$ um polinómio qualquer de grau $n \geq 1$ e*

$$\|q_n\| = \max_{a \leq x \leq b} |q_n(x)|.$$

Tem-se então a desigualdade

$$(15.33) \quad \|q_n\| \geq |c_n| \frac{(b-a)^n}{2^{2n-1}}.$$

Além disso a igualdade (15.33) verifica-se se

$$q_n(x) = c_n \frac{(b-a)^n}{2^{2n-1}} T_n\left(\frac{2x-a-b}{b-a}\right).$$

Demonstração. Consideremos a transformação

$$t = \frac{2x-a-b}{b-a}.$$

Ela aplica o intervalo $a \leq x \leq b$ de uma maneira biunívoca sobre o intervalo $-1 \leq t \leq 1$. Visto que

$$x = \frac{b-a}{2} t + \frac{b+a}{2}$$

tem-se

$$x^n = \left(\frac{b-a}{2}\right)^n t^n + \text{termos de grau inferior},$$

logo

$$q_n(x) = c_n \left(\frac{b-a}{2}\right)^n p_n(t),$$

sendo $p_n(t)$ um polinómio em t de grau n com o coeficiente de x^n igual a 1. Aplicando o teorema 15.11 a p_n obtemos o teorema 15.12.

15.17. Aplicação à fórmula de erro na interpolação

Voltamos agora à fórmula do erro (15.27) para a interpolação polinomial. Se se colhemos os pontos de interpolação x_0, x_1, \dots, x_n de modo que coincidam com o $n+1$ zeros do polinómio de Tchebycheff T_{n+1} , pode escrever-se (15.27) na forma

$$f(x) - P(x) = \frac{T_{n+1}(x)}{2^n(n+1)!} f^{(n+1)}(c).$$

Os pontos x_0, x_1, \dots, x_n estão todos no intervalo aberto $(-1, 1)$ e são definidos por

$$x_k = \cos\left(\frac{2k+1}{n+1} \frac{\pi}{2}\right) \quad \text{para } k = 0, 1, 2, \dots, n.$$

Se x pertence ao intervalo $[-1, 1]$ temos $|T_{n+1}(x)| \leq 1$ e o erro é estimado pela desigualdade

$$|f(x) - P(x)| \leq \frac{1}{2^n(n+1)!} |f^{(n+1)}(c)|.$$

Se a interpolação tem lugar num intervalo $[a, b]$ com os pontos

$$y_k = \frac{b-a}{2} x_k + \frac{b+a}{2}$$

como pontos de interpolação, o produto

$$A(x) = (x - y_0)(x - y_1) \cdots (x - y_n)$$

satisfaz à desigualdade $|A(x)| \leq (b - a)^{n+1}/2^{2n+1}$ para todo x em $[a, b]$. A correspondente estimativa para $f(x) - P(x)$ é

$$|f(x) - P(x)| \leq \frac{(b - a)^{n+1}}{2^{2n+1}(n + 1)!} |f^{(n+1)}(c)|.$$

15.18. Exercícios

Neste conjunto de exercícios T_n representa um polinómio de Tchebycheff de grau n .

1. Provar que $T_n(-x) = (-1)^n T_n(x)$. Isto prova que T_n é uma função par quando n é par e uma função ímpar quando n é ímpar.
2. (a) Provar que no intervalo aberto $(-1, 1)$ a derivada T'_n é dada por

$$T'_n(x) = \frac{n \operatorname{sen} n\theta}{\operatorname{sen} \theta}, \quad \text{com} \quad \theta = \arccos x.$$

(b) Calcular $T'_n(1)$ e $T'_n(-1)$.

3. Provar que $\int_0^x T_n(u) du = \frac{1}{2} \left\{ \frac{T_{n+1}(x) - T_{n+1}(0)}{n+1} - \frac{T_{n-1}(x) - T_{n-1}(0)}{n-1} \right\}$ se $n \geq 2$.

4. (a) Provar que $2T_m(x)T_n(x) = T_{m+n}(x) + T_{m-n}(x)$.

(b) Provar que $T_{mn}(x) = T_m(T_n(x)) = T_n(T_m(x))$.

5. Se $x = \cos \theta$, provar que $\operatorname{sen} \theta \operatorname{sen} n\theta$ é um polinómio em x , e determinar o seu grau.
6. O polinómio de Tchebycheff T_n satisfaz à equação diferencial

$$(1 - x^2)y'' - xy' + n^2y = 0$$

em todo o eixo real. Demonstra-lo por cada um dos métodos seguintes:

- (a) Derivando a relação $T'_n(x) \operatorname{sen} \theta = n \operatorname{sen} n\theta$ obtida no Exercício 2(a).
- (b) Introduzindo a mudança de variável $x = \cos \theta$ na equação diferencial

$$\frac{d^2(\cos n\theta)}{d\theta^2} = -n^2 \cos n\theta.$$

7. Determinar, em termos de polinómios de Tchebycheff, um polinómio $Q(x)$ de grau $\leq n$ que melhor aproxime x^{n+1} no intervalo $[-1, 1]$ relativamente à norma maximal.
8. Determinar um polinómio de melhor aproximação de grau ≤ 4 para a função $f(x) = x^5$ no intervalo $[0, 1]$, relativo à norma maximal.
9. Um polinómio P diz-se *primário* se o coeficiente do termo de maior grau é a unidade. Para um dado intervalo $[a, b]$ seja $\|P\|$ o máximo de $|P|$ em $[a, b]$. Demonstrar cada uma das seguintes proposições:
 - (a) Se $b - a < 4$, para todo $\epsilon > 0$ existe um polinómio primário P com $\|P\| < \epsilon$.
 - (b) Se para cada $\epsilon > 0$ existir um polinómio primário P com $\|P\| < \epsilon$, então $b - a < 4$. Por

outras palavras, polinómios primários com norma arbitrariamente pequena existem se, só se, o intervalo $[a, b]$ tem amplitude inferior a 4.

10. Os polinómios de Tchebycheff verificam as seguintes relações de ortogonalidade:

$$\int_{-1}^1 \frac{T_n(x)T_m(x)}{\sqrt{1-x^2}} dx = \begin{cases} 0 & \text{si } n \neq m, \\ \pi & \text{si } n = m = 0, \\ \frac{\pi}{2} & \text{si } n = m > 0. \end{cases}$$

Demonstrá-los por cada um dos seguintes métodos:

- (a) A partir da equação diferencial do Exercício 6 deduzir que

$$T_m(x) \frac{d}{dx} \left(\sqrt{1-x^2} T'_n(x) \right) + n^2 \frac{T_n(x)T_m(x)}{\sqrt{1-x^2}} = 0.$$

Escrever a fórmula correspondente com n e m permutados, subtrair as duas equações, integrá-las de -1 a $+1$.

- (b) Utilizar as relações de ortogonalidade

$$\int_0^\pi \cos m\theta \cos n\theta d\theta = \begin{cases} 0 & \text{se } n \neq m, \quad n > 0, \quad m > 0, \\ \pi & \text{se } n = m = 0, \\ \frac{\pi}{2} & \text{se } n = m > 0, \end{cases}$$

e introduzir a mudança de variável $x = \cos \theta$.

11. Provar que para $-1 < x < 1$ se tem

$$\frac{T_n(x)}{\sqrt{1-x^2}} = (-1)^n \frac{2^n n!}{(2n)!} \frac{d^n}{dx^n} (1-x^2)^{n-1/2}.$$

12. Sejam y_1, y_2, \dots, y_n , n números reais, e sejam

$$x_k = \cos \frac{(2k-1)\pi}{2n} \quad \text{para } k = 1, 2, \dots, n.$$

Seja P o polinómio de grau $\leq n-1$ que toma o valor y_k em x_k para $1 \leq k \leq n$. Se x não for um dos x_k provar que

$$P(x) = \frac{1}{n} \sum_{k=1}^n (-1)^{k-1} y_k \sqrt{1-x_k^2} \frac{T_n(x)}{x-x_k}.$$

13. Seja P um polinómio de grau $\leq n-1$ tal que

$$\sqrt{1-x^2} |P(x)| \leq 1$$

para $-1 \leq x \leq 1$. Provar que $\|P\| \leq n$, onde $\|P\|$ é o máximo de $|P|$ no intervalo $[-1, 1]$.

[Sugestão: Recorrer ao Exercício 12. Considerar três casos: $x_1 \leq x \leq 1$; $-1 \leq x \leq x_n$; $x_n \leq x \leq x_1$; nos dois primeiros fazer uso do Exercício 15(a) da Secção 15.9. No terceiro caso verificar que $\sqrt{1-x^2} \geq \sin(\pi/2n) > 1/n$].

Nos Exercícios 14 a 18, $U_n(x) = T'_{n+1}(x)/(n+1)$ para $n = 0, 1, 2, \dots$.

14. (a) Provar que $U_n(x) = 2xU_{n-1}(x) - U_{n-2}(x)$ para $n \geq 2$.

(b) Determinar a forma explícita dos polinómios U_0, U_1, \dots, U_5 .

(c) Provar que $|U_n(x)| \leq n+1$ se $-1 \leq x \leq 1$.

15. Mostrar que U_n satisfaz à equação diferencial

$$(1-x^2)y'' - 3xy' + n(n+2)y = 0.$$

16. Deduzir as relações de ortogonalidade

$$\int_{-1}^1 \sqrt{1-x^2} U_m(x) U_n(x) dx = \begin{cases} 0 & \text{se } m \neq n, \\ \frac{\pi}{2} & \text{se } m = n. \end{cases}$$

17. Provar que

$$\sqrt{1-x^2} U_n(x) = (-1)^n \frac{2^n(n+1)!}{(2n+1)!} \frac{d^n}{dx^n} (1-x^2)^{n+1/2}.$$

18. Sejam y_1, y_2, \dots, y_n , n números reais e

$$x_k = \cos \frac{k\pi}{n+1} \quad \text{para } k = 1, 2, \dots, n.$$

Seja P o polinómio de grau $\leq n-1$ que toma o valor y_k em x_k para $1 \leq k \leq n$. Se x não for um dos x_k provar que

$$P(x) = \frac{1}{n+1} \sum_{k=1}^n (-1)^{k-1} (1-x_k^2) y_k \frac{U_n(x)}{x-x_k}.$$

15.19. Integração aproximada. A regra trapezoidal

Muitos problemas, tanto de matemática pura como aplicada, conduzem-nos a novas funções cujas propriedades não foram estudadas ou cujos valores não foram tabulados. Para satisfazer certas necessidades práticas das ciências aplicadas, torna-se muitas vezes necessário obter informação quantitativa acerca de tais funções, quer na forma gráfica, quer na forma numérica. Muitas destas funções aparecem em integrais do tipo

$$F(x) = \int_a^x f(t) dt,$$

onde o integrando f é dado por uma fórmula explícita ou é conhecida em parte por valores tabulados. A parte restante deste capítulo é dedicada ao estudo de alguns dos métodos mais elementares para a determinação das aproximações numéricas de

tais integrais. A ideia base é muito simples. Aproximamos o integrando f por outra função P cujo integral seja de cálculo fácil, e tomamos então o integral de P como aproximação do integral f .

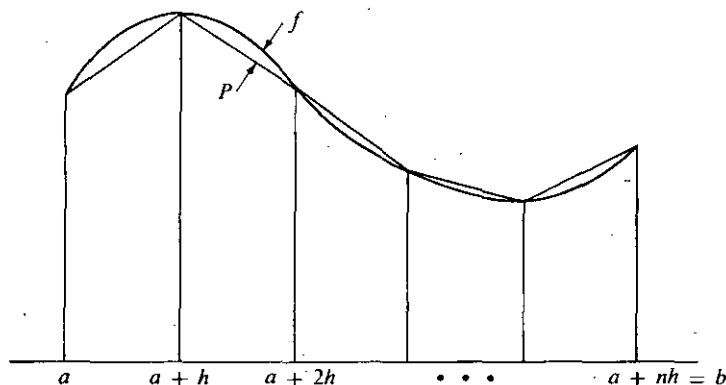


FIG. 15.3 A regra trapezoidal obtida mediante interpolação seccionalmente linear.

Se f é não negativa o integral $\int_a^b f(x)dx$ representa a área do conjunto de ordenadas de f relativo a $[a, b]$. Esta interpretação geométrica do integral sugere imediatamente certos procedimentos para a integração aproximada. A figura 15.3 mostra um exemplo de uma função f com valores conhecidos em $n+1$ pontos igualmente espaçados, $a, a+h, a+2h, \dots, a+nh=b$, sendo $h=(b-a)/n$. Seja $x_k = a+kh$. Para cada $k=0, 1, 2, \dots, n-1$ o gráfico de f referente ao intervalo $[x_k, x_{k+1}]$ foi aproximado por uma função linear que coincide com f nos extremos x_k e x_{k+1} . Seja P a correspondente função de interpolação seccionalmente linear definida em todo o intervalo $[a, b]$. Então tem-se

$$(15.34) \quad P(x) = \frac{x_{k+1} - x}{h} f(x_k) + \frac{x - x_k}{h} f(x_{k+1}) \quad \text{se } x_k \leq x \leq x_{k+1}.$$

Integrando no intervalo $[x_k, x_{k+1}]$ encontramos

$$\int_{x_k}^{x_{k+1}} P(x) dx = h \frac{f(x_k) + f(x_{k+1})}{2}.$$

Quando f é positiva esta é a área do trapézio determinado pelo gráfico de P relativo a $[x_k, x_{k+1}]$. A fórmula é válida, naturalmente, mesmo se f não é sempre positiva. Somando os integrais referentes a todos os subintervalos $[x_k, x_{k+1}]$ obtemos

$$\begin{aligned}
 (15.35) \quad \int_a^b P(x) dx &= \frac{h}{2} \sum_{k=0}^{n-1} [f(x_k) + f(x_{k+1})] \\
 &= \frac{h}{2} \left(f(a) + 2 \sum_{k=1}^{n-1} f(a + kh) + f(b) \right).
 \end{aligned}$$

Para se utilizar esta soma como uma aproximação do integral $\int_a^b f(x) dx$ necessitamos uma estimativa do erro $\int_a^b f(x) dx - \int_a^b P(x) dx$. Se f possui derivada de segunda ordem contínua em $[a, b]$ este erro é dado pelo seguinte teorema.

TEOREMA 15.13. REGRA TRAPEZOIDAL. Se f possui derivada de segunda ordem f'' contínua em $[a, b]$ e se n é um inteiro positivo e $h = (b - a)/n$, então tem-se

$$(15.36) \quad \int_a^b f(x) dx = \frac{b-a}{2n} \left(f(a) + 2 \sum_{k=1}^{n-1} f(a + kh) + f(b) \right) - \frac{(b-a)^3}{12n^2} f''(c)$$

para algum c em $[a, b]$.

Nota: A fórmula (15.36) é conhecida por *regra trapezoidal*. O termo $-f''(c)(b-a)^3/12n^2$ representa o erro na aproximação de $\int_a^b f(x) dx$ por $\int_a^b P(x) dx$. Desde que o valor máximo de f'' em $[a, b]$ seja conhecido podemos aproximar o integral de f até à precisão desejada fazendo n suficientemente grande. Observe-se que não é necessário qualquer conhecimento da função interpoladora P para utilizar esta fórmula. É unicamente necessário conhecer os valores de f nos pontos $a, a+h, \dots, a+nh$, e ter uma estimativa para $|f''(c)|$.

Demonstração. Seja P a função interpoladora dada por (15.34), onde $x_k = a + kh$. Em cada subintervalo $[x_k, x_{k+1}]$ aplicamos a estimativa do erro para a interpolação linear dada pelo teorema 15.3 e encontramos

$$(15.37) \quad f(x) - P(x) = (x - x_k)(x - x_{k+1}) \frac{f''(c_k)}{2!}$$

para certo c_k em (x_k, x_{k+1}) . Sejam M_2 e m_2 os valores máximos e mínimo, respectivamente, de f'' em $[a, b]$ e

$$B(x) = (x - x_k)(x_{k+1} - x)/2.$$

Então $B(x) \geq 0$ no intervalo $[x_k, x_{k+1}]$ e de (15.37) obtemos as desigualdades

$$m_2 B(x) \leq P(x) - f(x) \leq M_2 B(x)$$

neste intervalo. Integrando temos

$$(15.38) \quad m_2 \int_{x_k}^{x_{k+1}} B(x) dx \leq \int_{x_k}^{x_{k+1}} [P(x) - f(x)] dx \leq M_2 \int_{x_k}^{x_{k+1}} B(x) dx.$$

O integral de B é dado por

$$\int_{x_k}^{x_{k+h}} B(x) dx = \frac{1}{2} \int_{x_k}^{x_{k+h}} (x - x_k)(x_{k+1} - x) dx = \frac{1}{2} \int_0^h t(h-t) dt = \frac{h^3}{12}.$$

Por conseguinte as desigualdades (15.38) dão-nos

$$m_2 \leq \frac{12}{h^3} \int_{x_k}^{x_{k+h}} [P(x) - f(x)] dx \leq M_2.$$

Somando estas desigualdades para $k = 0, 1, 2, \dots, n-1$ e dividindo por n , obtemos

$$m_2 \leq \frac{12}{nh^3} \int_a^b [P(x) - f(x)] dx \leq M_2.$$

Porque a função f'' é contínua em $[a, b]$, toma qualquer valor entre o mínimo m_2 e o M_2 em algum ponto de $[a, b]$. Em particular, temos

$$f''(c) = \frac{12}{nh^3} \int_a^b [P(x) - f(x)] dx$$

para algum c em $[a, b]$. Por outras palavras,

$$\int_a^b f(x) dx = \int_a^b P(x) dx - \frac{nh^3}{12} f''(c).$$

Com (15.35) e a relação $h = (b-a)/n$ obtemos (15.36).

Para deduzir a regra trapezoidal usámos um polinómio linear para interpolar entre cada par de valores consecutivos de f . Interpolando com polinómios de grau superior podem obter-se fórmulas mais rigorosas. Consideremos a seguir um caso particular importante que é notável pela sua simplicidade e precisão.

15.20. Regra de Simpson

A curva a traço contínuo da figura 15.4 é o gráfico de uma função f correspondente ao intervalo $[a, b]$. O ponto médio do intervalo, $(a+b)/2$, designa-se por m . A curva a tracejado é o gráfico de um polinómio P do segundo grau, cujos valores coincidem com os de f nos três pontos a, m e b . Se usarmos o integral $\int_a^b P(x) dx$ como uma aproximação de $\int_a^b f(x) dx$ somos conduzidos a uma fórmula de integração aproximada conhecida por *regra de Simpson*.

Em vez de determinarmos P explicitamente, introduzimos uma transformação linear que transforma o intervalo $[a, b]$ no intervalo $[0, 2]$. Se escrevermos

$$t = \frac{x-a}{m-a}, \quad \text{ou} \quad x = a + (m-a)t,$$

vemmos que t toma os valores 0, 1, 2 quando x toma os valores a, m, b . Seja agora

$$\varphi(t) = P[a + (m - a)t].$$

Então $\varphi(t)$ é um polinômio quadrático em t que toma os valores $P(a)$, $P(m)$, $P(b)$ nos pontos $t = 0, 1, 2$, respectivamente. Também se tem

$$\int_0^2 \varphi(t) dt = \int_0^2 P[a + (m - a)t] dt = \frac{1}{m - a} \int_a^b P(x) dx;$$

logo

$$(15.39) \quad \int_a^b P(x) dx = (m - a) \int_0^2 \varphi(t) dt = \frac{b - a}{2} \int_0^2 \varphi(t) dt.$$

Utilizemos agora a fórmula de interpolação de Newton para construirmos φ . Temos

$$\varphi(t) = \varphi(0) + t \Delta \varphi(0) + t(t - 1) \frac{\Delta^2 \varphi(0)}{2!},$$

com $\Delta \varphi(t) = \varphi(t + 1) - \varphi(t)$. Integrando de 0 a 2 obtemos

$$\int_0^2 \varphi(t) dt = 2\varphi(0) + 2\Delta \varphi(0) + \frac{1}{3}\Delta^2 \varphi(0).$$

Porque $\Delta \varphi(0) = \varphi(1) - \varphi(0)$ e $\Delta^2 \varphi(0) = \varphi(2) - 2\varphi(1) + \varphi(0)$, o integral é igual a

$$\int_0^2 \varphi(t) dt = \frac{1}{3}[\varphi(0) + 4\varphi(1) + \varphi(2)] = \frac{1}{3}[P(a) + 4P(m) + P(b)].$$

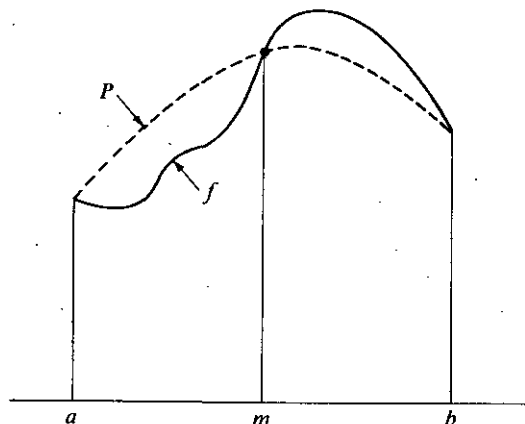


FIG. 15.4. Interpolação por um polinômio quadrático P .

Usando (15.39) e o facto de que P coincide com f em a, m, b , obtemos

$$(15.40) \quad \int_a^b P(x) dx = \frac{b-a}{6} [f(a) + 4f(m) + f(b)].$$

Portanto, podemos escrever

$$\int_a^b f(x) dx = \frac{b-a}{6} [f(a) + 4f(m) + f(b)] + R,$$

sendo $R = \int_a^b f(x) dx - \int_a^b P(x) dx$.

Se f é um polinómio do segundo grau, P é idêntico a f e o erro R é zero. É um facto notável o ter-se ainda $R=0$ quando f é um polinómio cúbico. Para demonstrar esta propriedade usamos a estimativa do erro para a interpolação de Lagrange dada pelo teorema 15.3 e escrevemos

$$(15.41) \quad f(x) - P(x) = (x-a)(x-m)(x-b) \frac{f'''(c)}{3!},$$

com $c \in (a, b)$. Quando f é um polinómio cúbico, a terceira derivada f''' é constante, digamos $f'''(x) = C$, e a fórmula precedente vem

$$f(x) - P(x) = \frac{C}{6} (x-a)(x-m)(x-b) = \frac{C}{6} (t+h)t(t-h),$$

com $t = x - m$ e $h = (b-a)/2$. Deste modo

$$R = \int_a^b [f(x) - P(x)] dx = \frac{C}{6} \int_{-h}^h (t^3 - h^2 t) dt = 0,$$

já que o último integrando é uma função ímpar. Esta propriedade está ilustrada na figura 15.5. A curva a tracejado é o gráfico de um polinómio cúbico f que coincide com P em a, m e b . Neste caso $R = \int_a^b [f(x) - P(x)] dx = A_1 - A_2$, onde A_1 e A_2 são áreas das duas regiões sombreadas. Porque $R=0$ as duas regiões têm áreas iguais.

Acabámos de ver que (15.40) é válida se P é um polinómio de grau ≤ 3 que coincide com f em a, m e b . Escolhendo cuidadosamente este polinómio podemos melhorar consideravelmente a estimativa do erro (15.41). Já impusemos três condições a P , a saber, $P(a)=f(a)$, $P(m)=f(m)$, $P(b)=f(b)$. Impomos agora uma quarta condição $P'(m)=f'(m)$. Isto significará que P e f têm os mesmos declives em $(m, f(m))$, e será de esperar que isto melhore a aproximação de f por P no intervalo $[a, b]$.

Para provarmos que pode sempre escolher-se um tal P , designemos por Q o polinómio quadrático que coincide com f em a, m, b e seja

$$P(x) = Q(x) + A(x-a)(x-m)(x-b),$$

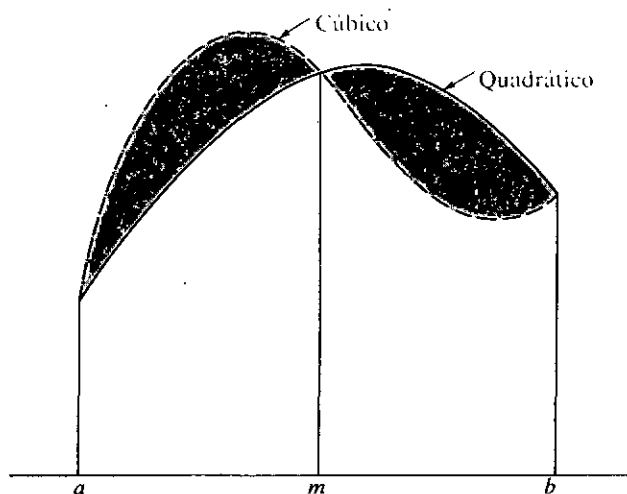


FIG. 15.5. As duas regiões sombreadas têm áreas iguais para qualquer polinômio interpolador do terceiro grau.

sendo A um valor constante a determinar. Para A qualquer, este polinômio cúbico P coincide com Q e portanto com f , em a , m e b . Escolhamos agora A de modo que $P'(m) = f'(m)$. Derivando ambos os membros e fazendo $x = m$ obtemos

$$P'(m) = Q'(m) + A(m - a)(m - b).$$

Por conseguinte se tomamos $A = [f'(m) - Q'(m)] / [(m - a)(m - b)]$ satisfaz-se também a condição $P'(m) = f'(m)$.

Mostramos a seguir que para esta escolha de P se tem

$$(15.42) \quad f(x) - P(x) = (x - a)(x - m)^2(x - b) \frac{f^{(4)}(z)}{4!}$$

para algum z em (a, b) , com tanto que a derivada $f^{(4)}$ exista em $[a, b]$. Para demonstrar (15.42) raciocinamos como na demonstração do teorema 15.3. Observamos primeiro que é trivial a verificação de (15.42) para qualquer z se $x = a$, m ou b . Portanto, admitimos $x \neq a$, $x \neq m$, $x \neq b$, fixamos x , e introduzimos uma nova função F definida em $[a, b]$ por

$$F(t) = A(x)[f(t) - P(t)] - A(t)[f(x) - P(x)],$$

com

$$A(t) = (t - a)(t - m)^2(t - b).$$

Observe-se que $F(t) = 0$ para $t = a$, m , b e x . Pelo teorema de Rolle, $F'(t)$ anula-se em cada um dos três intervalos abertos determinados por estes quatro pontos. Em com

plemento, $F'(m) = 0$ porque $A'(m) = 0$ e $f'(m) = P'(m)$. Deste modo $F'(t) = 0$ para pelo menos quatro pontos distintos de (a, b) . Pelo teorema de Rolle $F''(t) = 0$ em pelo menos três pontos, $F'''(t) = 0$ em pelo menos dois pontos, e $F^{(4)}(t) = 0$ em pelo menos um ponto, seja $t = z$. Da definição de F encontramos

$$\begin{aligned} F^{(4)}(t) &= A(x)[f^{(4)}(t) - P^{(4)}(t)] - A^{(4)}(t)[f(x) - P(x)] \\ &= A(x)f^{(4)}(t) - 4! [f(x) - P(x)]. \end{aligned}$$

Quando fazemos $t = z$ nesta equação obtemos (15.42).

É agora uma questão simples provar a regra de Simpson na forma seguinte:

TEOREMA 15.14. REGRA DE SIMPSON. *Se f tem derivada de quarta ordem contínua em $[a, b]$ e se $m = (a + b)/2$, então*

$$(15.43) \quad \int_a^b f(x) dx = \frac{b-a}{6} [f(a) + 4f(m) + f(b)] - \frac{(b-a)^5}{2880} f^{(4)}(c)$$

para algum c em $[a, b]$.

Demonstração. Sejam M_4 e m_4 respectivamente o máximo e o mínimo de $f^{(4)}$ em $[a, b]$ e $B(x) = -(x-a)(x-m)^2(x-b)/4!$. Porque $B(x) \geq 0$ para todo x em $[a, b]$, a equação (15.42) conduz-nos às desigualdades

$$m_4 B(x) \leq P(x) - f(x) \leq M_4 B(x).$$

Integrando, encontramos

$$(15.44) \quad m_4 \int_a^b B(x) dx \leq \int_a^b [P(x) - f(x)] dx \leq M_4 \int_a^b B(x) dx.$$

Para calcular o integral $\int_a^b B(x) dx$, fazemos $h = (b - a)/2$ e temos

$$\begin{aligned} \int_a^b B(x) dx &= -\frac{1}{4!} \int_a^b (x-a)(x-m)^2(x-b) dx = -\frac{1}{4!} \int_{-h}^h (t+h)t^2(t-h) dt \\ &= -\frac{2}{4!} \int_0^h t^2(t^2 - h^2) dt = \frac{1}{4!} \frac{4h^5}{15} = \frac{(b-a)^5}{2880}. \end{aligned}$$

Portanto as desigualdades em (15.44) dão-nos

$$m_4 \leq \frac{2880}{(b-a)^5} \int_a^b [P(x) - f(x)] dx \leq M_4.$$

Mas visto que $f^{(4)}$ é contínua em $[a, b]$, toma todo e qualquer valor entre o mínimo m_4 e o máximo M_4 algures em $[a, b]$. Assim

$$f^{(4)}(c) = \frac{2880}{(b-a)^5} \int_a^b [P(x) - f(x)] dx$$

para algum c em $[a, b]$. Porque $\int_a^b P(x) dx = \frac{1}{6}(b-a)[f(a) + 4f(m) + f(b)]$, esta equação dá-nos (15.43).

A regra de Simpson é de interesse particular porque a sua precisão é superior ao que poderia esperar-se a partir do conhecimento da função f simplesmente em três pontos. Se os valores de f se conhecem num número ímpar de pontos igualmente espaçados, $a, a+h, \dots, a+2nh$, habitualmente é mais simples aplicar a regra de Simpson a cada um dos intervalos $[a, a+2h]$, $[a+2h, a+4h]$, ... do que usar um polinómio interpolador de grau $\leq 2n$ sobre o intervalo completo $[a, a+2nh]$. Aplicando a regra de Simpson desta maneira, obtemos a seguinte extensão do teorema 15.14.

TEOREMA 15.15. RÉGRA DE SIMPSON GENERALIZADA. *Admita-se que f tem uma derivada de quarta ordem contínua em $[a, b]$. Seja $h = (b-a)/(2n)$ e $f_k = f(a+kh)$ para $k = 1, 2, \dots, 2n-1$. Então tem-se*

$$\int_a^b f(x) dx = \frac{b-a}{6n} \left(f(a) + 4 \sum_{k=1}^n f_{2k-1} + 2 \sum_{k=1}^{n-1} f_{2k} + f(b) \right) - \frac{(b-a)^5}{2880n^4} f^{(4)}(\bar{c})$$

para algum \bar{c} em $[a, b]$.

A demonstração deste teorema é solicitada no Exercício 9 da secção seguinte.

15.21. Exercícios

1. (a) Aplicar a regra trapezoidal, com $n = 10$, para calcular o valor do integral

$$\log 2 = \int_1^2 \frac{dx}{x}.$$

Obter limites superior e inferior do erro. (Ver Exercício 10 b) para comparar a precisão com a que se obtém com a regra de Simpson).

(b) Qual é o menor valor de n que pode garantir seis casas decimais exactas no cálculo de $\log 2$ por este método?

2. (a) Provar que existe um número positivo c no intervalo $[0, 1]$ tal que a fórmula

$$\int_{-1}^1 f(x) dx = f(c) + f(-c)$$

é exacta para todos os polinómios de grau ≤ 3 .

(b) Generalizar o resultado da alínea (a) para um intervalo arbitrário, isto é, provar que existem constantes c_1 e c_2 em $[a, b]$ tais que a fórmula

$$\int_a^b f(x) dx = \frac{b-a}{2} [f(c_1) + f(c_2)]$$

é exacta para todos os polinómios de grau ≤ 3 . Expressar c_1 e c_2 em função de a e b .

3. (a) Provar que existe uma constante positiva c tal que a fórmula

$$\int_{-1/2}^{1/2} f(x) dx = \frac{1}{3} [f(-c) + f(0) + f(c)]$$

é exacta para todos os polinómios de grau ≤ 3 .

(b) Generalizar o resultado da alínea (a) para um intervalo arbitrário, isto é, provar que existem constantes c_1 e c_2 em $[a, b]$ tais que a fórmula

$$\int_a^b f(x) dx = \frac{b-a}{3} \left[f(c_1) + f\left(\frac{a+b}{2}\right) + f(c_2) \right]$$

é exacta para todos os polinómios de grau ≤ 3 . Expressar c_1 e c_2 em função de a e b .

4. Demonstrar que existem constantes positivas a e b tais que a fórmula

$$\int_0^\infty e^{-x^2} f(x) dx = \frac{1}{4} [af(b) + bf(a)]$$

é exacta para todos os polinómios de grau ≤ 3 .

5. Demonstrar que existe uma constante positiva c tal que a fórmula

$$\int_{-\infty}^\infty e^{-x^2} f(x) dx = \frac{\sqrt{\pi}}{6} [f(-c) + 4f(0) + f(c)]$$

é exacta para todos os polinómios de grau ≤ 5 .

6. Seja P_n o polinómio interpolador de grau $\leq n$ que coincide com f em $n+1$ pontos distintos x_0, x_1, \dots, x_n .

(a) Mostrar que existem constantes $A_0(n), \dots, A_n(n)$, dependendo unicamente dos números x_0, x_1, \dots, x_n , a e b , e não de f , tais que

$$\int_a^b P_n(x) dx = \sum_{k=0}^n A_k(n) f(x_k).$$

Os números $A_k(n)$ chamam-se *pesos*. (Por vezes designam-se também por números de Christoffel).

(b) Para um dado conjunto de pontos de interpolação distintos e um dado intervalo $[a, b]$, sejam $W_0(n), W_1(n), \dots, W_n(n)$, $n+1$ constantes tais que a fórmula

$$\int_a^b f(x) dx = \sum_{k=0}^n W_k(n) f(x_k)$$

é exacta para todos os polinómios de grau $\leq n$. Provar que

$$\sum_{k=0}^n x_k^r W_k(n) = \frac{b^{r+1} - a^{r+1}}{r+1} \quad \text{para } r = 0, 1, \dots, n.$$

Este é um sistema de $n+1$ equações lineares que pode ser utilizado para determinar os pesos. Pode provar-se que este sistema tem sempre uma solução que é única. Pode também provar-se que por uma escolha adequada de pontos de interpolação é possível fazer todos os pesos iguais. Quando os pesos são todos iguais a fórmula diz-se de integração de Tchebycheff. Os Exercícios 2 e 3 são exemplos de fórmulas de integração de Tchebycheff. O exercício que se segue mostra que para uma escolha adequada dos pontos de interpolação, a fórmula de integração resultante é exacta para todos os polinómios de grau $\leq 2n+1$.

7. Neste exercício podem usar-se as propriedades dos polinómios de Legendre estabelecidas nos Secções 6.19 e 6.20. Sejam x_0, x_1, \dots, x_n os zeros do polinómio de Legendre $P_{n+1}(x)$. Esses zeros são distintos e estão no intervalo $[-1, 1]$. Seja $f(x)$ qualquer polinómio em x de grau $\leq 2n+1$. Divida-se $f(x)$ por $P_{n+1}(x)$ e escreva-se

$$f(x) = P_{n+1}(x)Q(x) + R(x),$$

onde os polinómios Q e R têm grau $\leq n$.

- (a) Provar que o polinómio R coincide com f nos zeros de P_{n+1} e que

$$\int_{-1}^1 f(x) dx = \int_{-1}^1 R(x) dx.$$

- (b) Provar que existem $n+1$ pesos $W_0(n), \dots, W_n(n)$ (independentes de f) tais que

$$\int_{-1}^1 f(x) dx = \sum_{k=0}^n W_k(n) f(x_k).$$

Isto dá-nos uma fórmula de integração com $n+1$ pontos de interpolação que é exacta para todos os polinómios de grau $\leq 2n+1$.

- (c) Fazer $n=2$ e mostrar que a fórmula da alínea (b) vem

$$\int_{-1}^1 f(x) dx = \frac{8}{9} f(-\sqrt{\frac{3}{5}}) + \frac{8}{9} f(0) + \frac{8}{9} f(\sqrt{\frac{3}{5}}).$$

Esta fórmula é exacta para todos os polinómios de grau ≤ 5 .

- (d) Introduzir uma adequada transformação linear e escrever de novo a fórmula da alínea (c) para um intervalo $[a, b]$.

8. Este exercício descreve um método de Peano para deduzir a fórmula que dá o erro na regra de Simpson.

- (a) Usar repetidas vezes o método de integração por partes para deduzir a relação

$$\int u(t)v'''(t) dt = u(t)v''(t) - u'(t)v'(t) + u''(t)v(t) - \int g(t) dt,$$

com $g(t) = u'''(t)v(t)$.

- (b) Supor que φ tem uma derivada de quarta ordem contínua do intervalo $[-1, 1]$. Tomar

$$v(t) = t(1-t)^2/6, \quad u(t) = \varphi(t) + \varphi(-t),$$

e usar a alínea (a) para provar que

$$\int_{-1}^1 \varphi(t) dt = \frac{1}{3} [\varphi(-1) + 4\varphi(0) + \varphi(1)] - \int_0^1 g(t) dt.$$

Mostrar depois que $\int_0^1 g(t) dt = \varphi^{(4)}(c)/90$ para algum c em $[-1, 1]$.

- (c) Introduzir uma transformação linear adequada para deduzir o teorema 15.14 do resultado da alínea (b).

9. (a) Sejam a_1, a_2, \dots, a_n números não negativos cuja soma é 1. Supor φ contínua num intervalo $[a, b]$. Se c_1, c_2, \dots, c_n são n pontos quaisquer em $[a, b]$ (não necessariamente distintos), provar que existe pelo menos um ponto c em $[a, b]$ tal que

$$\sum_{k=1}^n a_k \varphi(c_k) = \varphi(c).$$

[Sugestão: Sejam M e m o máximo e o mínimo de φ em $[a, b]$ e utilizemos a desigualdade $m \leq \varphi(c_k) \leq M$.]

(b) Usar a alínea (a) e o teorema 15.14 para deduzir a regra de Simpson generalizada do teorema 15.15.

10. Calcular $\log 2$ a partir da fórmula $\log 2 = \int_1^2 x^{-1} dx$ considerando a regra de Simpson generalizada com (a) $n = 2$; (b) $n = 5$. Dar, para cada caso, limites superior e inferior do erro.
11. (a) Seja $\varphi(t)$ um polinômio em t de grau ≤ 3 . Expressar $\varphi(t)$ mediante a fórmula de interpolação de Newton e integrar para deduzir a fórmula

$$\int_0^3 \varphi(t) dt = \frac{3}{8} [\varphi(0) + 3\varphi(1) + 3\varphi(2) + \varphi(3)].$$

(b) Seja P o polinômio interpolador de grau ≤ 3 que coincide com f nos pontos $a, a+h, a+2h, a+3h$, com $h > 0$. Utilizar a alínea (a) para demonstrar que

$$\int_a^{a+3h} P(x) dx = \frac{3h}{8} [f(a) + 3f(a+h) + 3f(a+2h) + f(a+3h)].$$

(c) Supor que f tem uma derivada contínua de quarta ordem em $[a, b]$, e seja $h = (b-a)/3$. Provar que

$$\int_a^b f(x) dx = \frac{b-a}{8} [f(a) + 3f(a+h) + 3f(a+2h) + f(b)] - \frac{(b-a)^5}{6480} f^{(4)}(c)$$

para algum c em $[a, b]$. Esta fórmula de integração aproximada chama-se a *regra de Cotes*.

(d) Usar a regra de Cotes para calcular $\log 2 = \int_1^2 x^{-1} dx$ e determinar limites superior e inferior para o erro.

12. (a) Usar a equação vectorial $r(t) = a \sin t \mathbf{i} + b \cos t \mathbf{j}$, onde $0 < b < a$, para mostrar que o perímetro L de uma elipse é dado por

$$L = 4a \int_0^{\pi/2} \sqrt{1 - k^2 \sin^2 t} dt,$$

com $k = \sqrt{a^2 - b^2}/a$.

(b) Provar que a regra de Simpson dá a fórmula

$$L = \frac{\pi}{3} [a + b + \sqrt{8(a^2 + b^2)}] - \frac{a\pi^5}{23040} f^{(4)}(c)$$

para um certo c em $[0, \pi/2]$, sendo $f(t) = \sqrt{1 - k^2 \sin^2 t}$.

15.22. A fórmula de somação de Euler

Seja n um inteiro positivo. Quando a regra trapezoidal (teorema 15.13) se aplica ao intervalo $[0, n]$ toma a forma

$$\int_0^n f(x) dx = \sum_{k=0}^{n-1} f(k) + \frac{1}{2}(f(n) - f(0)) - \frac{f''(c)n}{12}$$

para algum c em $[0, n]$. Se f é um polinómio quadrático então f'' é constante e portanto $f''(c) = f''(0)$. Neste caso a fórmula pode voltar a escrever-se na forma

$$(15.45) \quad \sum_{k=0}^n f(k) = \int_0^n f(x) dx + \frac{f(0) + f(n)}{2} + \frac{f''(0)n}{12}.$$

Esta fórmula é exacta quando f é um qualquer polinómio de grau ≤ 2 .

Euler estabeleceu uma generalização notável desta fórmula que é exacta para qualquer função com derivada de primeira ordem contínua. Pode usar-se para aproximar integrais por somas ou, o que é mais frequente, para calcular ou estimar somas por meio de integrais. Por esta razão é usualmente designada por fórmula de "somação" em vez de fórmula de integração. Pode estabelecer-se do modo seguinte:

TEOREMA 15.16. FÓRMULA DE SOMAÇÃO DE EULER. *Se f tem derivada de primeira ordem contínua em $[0, n]$, então tem-se*

$$(15.46) \quad \sum_{k=0}^n f(k) = \int_0^n f(x) dx + \frac{f(0) + f(n)}{2} + \int_0^n (x - [x] - \tfrac{1}{2}) f'(x) dx,$$

onde $[x]$ é o maior inteiro $\leq x$.

Demonstração. A integração por partes dá-nos

$$(15.47) \quad \int_0^n (x - \tfrac{1}{2}) f'(x) dx = (n - \tfrac{1}{2}) f(n) + \tfrac{1}{2} f(0) - \int_0^n f(x) dx.$$

Consideremos agora o integral $\int_0^n [x] f'(x) dx$ e escrevamo-lo como uma soma de integrais em cada um dos quais $[x]$ tem um valor fixo. Assim temos

$$\begin{aligned} \int_0^n [x] f'(x) dx &= \sum_{r=0}^{n-1} \int_r^{r+1} [x] f'(x) dx = \sum_{r=0}^{n-1} r \int_r^{r+1} f'(x) dx \\ &= \sum_{r=0}^{n-1} r(f(r+1) - f(r)) = \sum_{r=0}^{n-1} r f(r+1) - \sum_{r=0}^{n-1} r f(r) \\ &= - \sum_{r=0}^{n-1} f(r+1) + \sum_{r=0}^{n-1} (r+1) f(r+1) - \sum_{r=0}^{n-1} r f(r) \\ &= - \sum_{k=1}^n f(k) + n f(n) = - \sum_{k=0}^n f(k) + f(0) + n f(n). \end{aligned}$$

Subtraindo esta da equação (15.47), obtemos

$$\int_0^n (x - [x] - \tfrac{1}{2}) f'(x) dx = \sum_{k=0}^n f(k) - \frac{f(0) + f(n)}{2} - \int_0^n f(x) dx,$$

a qual é equivalente a (15.46).

O último integral do segundo membro de (15.46) pode escrever-se

$$\int_0^n (x - [x] - \tfrac{1}{2}) f'(x) dx = \int_0^n \varphi_1(x) f'(x) dx,$$

onde φ_1 é a função definida por

$$\varphi_1(x) = \begin{cases} x - [x] - \tfrac{1}{2} & \text{se } x \text{ não é inteiro,} \\ 0 & \text{se } x \text{ é inteiro.} \end{cases}$$

O gráfico de φ_1 está traçado na figura 15.6(a). Observe-se que $\varphi_1(x+1) = \varphi_1(x)$, o que significa que φ_1 é periódica de período 1. Também, se $0 < x < 1$ temos $\varphi_1(x) = x - \tfrac{1}{2}$ pelo que $\int_0^1 \varphi_1(t) dt = 0$.

A figura 15.6(b) mostra o gráfico de φ_2 , o integral indefinido de φ_1 , dado por

$$\varphi_2(x) = \int_0^x \varphi_1(t) dt.$$

É facilmente verificável que φ_2 é também periódica com período 1. Além disso, tem-se

$$\varphi_2(x) = \frac{x(x-1)}{2} \quad \text{se } 0 \leq x \leq 1.$$

Isto mostra que $-\tfrac{1}{8} \leq \varphi_2(x) \leq 0$ para todo x . As desigualdades $-\tfrac{1}{8} < \varphi_2(x) < 0$ são válidas excepto quando x é um inteiro ou metade de um inteiro.

O teorema que apresentamos a seguir dá-nos outra versão da fórmula de somação de Euler em função de φ_2 .

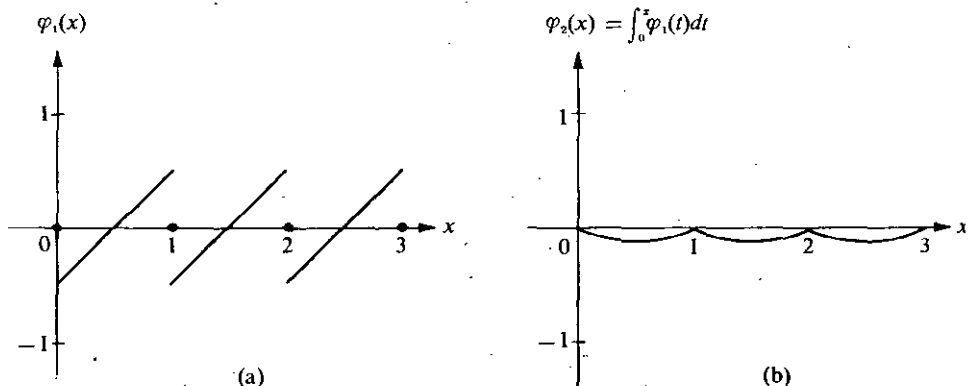


FIG. 15.6 Gráficos das funções periódicas φ_1 e φ_2 .

TEOREMA 15.17. Se f'' é contínua em $[0, n]$ tem-se

$$(15.48) \quad \sum_{k=0}^n f(k) = \int_0^n f(x) dx + \frac{f(0) + f(n)}{2} - \int_0^n \varphi_2(x) f''(x) dx.$$

Demonstração. Visto que $\varphi_2'(x) = \varphi_1(x)$ nos pontos em que φ_1 é contínua, temos

$$\int_0^n \varphi_1(x) f'(x) dx = \int_0^n \varphi_2'(x) f'(x) dx.$$

A integração por partes dá-nos

$$\int_0^n \varphi_2'(x) f'(x) dx = \varphi_2(x) f'(x) \Big|_0^n - \int_0^n \varphi_2(x) f''(x) dx = - \int_0^n \varphi_2(x) f''(x) dx,$$

porque $\varphi_2(n) = \varphi_2(0) = 0$. Aplicando esta fórmula em (15.46) obtemos (15.48).

Nota: Embora os teoremas 15.16 e 15.17 se refiram ao intervalo $[0, n]$, ambas as fórmulas são válidas quando se substitui 0 por 1 ou por qualquer inteiro positivo $< n$.

Para ilustrar o uso da fórmula de somação de Euler vamos deduzir a fórmula seguinte para $\log n!$.

TEOREMA 15.18. *Para qualquer inteiro e positivo n tem-se*

$$(15.49) \quad \log n! = (n + \tfrac{1}{2}) \log n - n + C + E(n),$$

onde $0 < E(n) < 1/(8n)$ e $C = 1 + \int_1^\infty t^{-2} \varphi_2(t) dt$.

Demonstração. Façamos $f(x) = \log x$ e apliquemos o teorema 15.17 ao intervalo $[a, n]$. Vem então

$$\sum_{k=1}^n \log k = \int_1^n \log x dx + \tfrac{1}{2} \log n + \int_1^n \frac{\varphi_2(x)}{x^2} dx.$$

Usando a relação $\int \log t dt = t \log t - t$ damos à igualdade anterior a forma

$$(15.50) \quad \log n! = (n + \tfrac{1}{2}) \log n - n + 1 + \int_1^n \frac{\varphi_2(t)}{t^2} dt.$$

Visto ser $|\varphi_2(t)| \leq \frac{1}{8}$ o integral impróprio $\int_2^\infty t^{-2} \varphi_2(t) dt$ converge absolutamente e podemos escrever

$$\int_1^n \frac{\varphi_2(t)}{t^2} dt = \int_1^\infty \frac{\varphi_2(t)}{t^2} dt - \int_n^\infty \frac{\varphi_2(t)}{t^2} dt.$$

Deste modo a equação (15.50) vem

$$\log n! = (n + \tfrac{1}{2}) \log n - n + C - \int_n^\infty \frac{\varphi_2(t)}{t^2} dt,$$

onde $C = 1 + \int_1^\infty t^{-2} \varphi_2(t) dt$. Visto que $-\frac{1}{8} < \varphi_2(t) < 0$ excepto quando t é um inteiro ou metade de um inteiro, obtemos

$$0 < - \int_n^{\infty} \frac{\varphi_2(t)}{t^2} dt < \frac{1}{8n},$$

Está assim provada (15.49), com $E(n) = - \int_n^{\infty} t^{-2} \varphi_2(t) dt$.

Do teorema 15.18 podemos derivar a fórmula de Stirling que permite uma estimativa de $n!$

TEOREMA 15.19. FÓRMULA DE STIRLING. *Se n é um inteiro positivo tem-se*

$$\sqrt{2\pi} n^{n+1/2} e^{-n} < n! < \sqrt{2\pi} n^{n+1/2} e^{-n} \left(1 + \frac{1}{4n}\right).$$

Demonstração. Utilizando a equação (15.49) e as desigualdades para $E(n)$ obtemos

$$\exp\left((n + \frac{1}{2}) \log n - n + C\right) < n! < \exp\left((n + \frac{1}{2}) \log n - n + C + \frac{1}{8n}\right),$$

onde $\exp(t) = e^t$. Tendo em conta que $e^x < 1 + 2x$, com $x = 1/(8n)$, podemos dar às desigualdades a forma

$$(15.51) \quad A n^{n+1/2} e^{-n} < n! < A n^{n+1/2} e^{-n} \left(1 + \frac{1}{4n}\right),$$

onde $A = e^C$. Para completar a demonstração necessitamos provar que $A = \sqrt{2\pi}$.

Deduzimos isto da desigualdade

$$(15.52) \quad \pi n \leq \left(\frac{2^{2n}(n!)^2}{(2n)!}\right)^2 \leq \frac{\pi(2n+1)}{2},$$

descoberta por John Wallis (1616-1703). Em primeiro lugar mostramos como a desigualdade de Wallis implica $A = \sqrt{2\pi}$; em seguida analisaremos a demonstração de (15.52).

Se fizermos

$$A_n = \frac{n!}{n^{n+1/2} e^{-n}}$$

a desigualdade em (15.51) implica

$$A < A_n < A \left(1 + \frac{1}{4n}\right).$$

Isto mostra que $A_n \rightarrow A$ quando $n \rightarrow \infty$. Em (15.52) escrevemos $n! = n^{n+1/2} e^{-n} A_n$ para obtermos

$$\pi n \leq \left(\frac{2^{2n} n^{2n+1} e^{-2n} A_n^2}{(2n)^{2n+1/2} e^{-2n} A_{2n}}\right)^2 \leq \frac{\pi(2n+1)}{2},$$

a qual é equivalente a

$$\pi \leq \frac{A_n^4}{2A_{2n}^2} \leq \pi \frac{2n+1}{2n}.$$

Façamos $n \rightarrow \infty$ nesta última desigualdade. Porque $A = e^C > 0$ obtemos

$$\pi \leq \frac{A^4}{2A^2} \leq \pi.$$

o que prova que $A^2 = 2\pi$, pelo que $A = \sqrt{2\pi}$ comó afirmámos.

Falta demonstrar a desigualdade de Wallis (15.52). Com esse objectivo introduzimos os números

$$I_n = \int_0^{\pi/2} \sin^n t \, dt,$$

sendo n um inteiro qualquer não negativo. Observe-se que $I_0 = \pi/2$ e $I_1 = 1$. Para $0 \leq t \leq \pi/2$ tem-se $0 \leq \sin t \leq 1$; logo $0 \leq \sin^{n+1} t \leq \sin^n t$. Isto mostra que a sucessão $\{I_n\}$ é monótona decrescente. Deste modo podemos escrever

$$(15.53) \quad \frac{1}{I_{2n} I_{2n-1}} \leq \frac{1}{I_{2n}^2} \leq \frac{1}{I_{2n} I_{2n+1}}.$$

Calcularemos agora cada membro desta desigualdade, isso conduzir-nos-á de imediato à desigualdade de Wallis.

A integração da identidade

$$\frac{d}{dt} (\cos t \sin^{n+1} t) = (n+1) \sin^n t - (n+2) \sin^{n+2} t$$

no intervalo $[0, \pi/2]$ dá-nos

$$0 = (n+1)I_n - (n+2)I_{n+2},$$

ou

$$(15.54) \quad I_{n+2} = \frac{n+1}{n+2} I_n.$$

Utilizando esta fórmula com n substituído por $2k-2$ encontramos

$$\frac{I_{2k}}{I_{2k-2}} = \frac{2k-1}{2k} = \frac{2k(2k-1)}{(2k)^2}.$$

Multiplicando estas equações para $k = 1, 2, \dots, n$, encontramos

$$\prod_{k=1}^n \frac{I_{2k}}{I_{2k-2}} = \prod_{k=1}^n \frac{2k(2k-1)}{(2k)^2} = \frac{(2n)!}{2^{2n}(n!)^2}.$$

O primeiro membro significa-se para I_{2n}/I_0 . Visto ser $I_0 = \pi/2$ obtemos

$$(15.55) \quad I_{2n} = \frac{(2n)!}{2^{2n}(n!)^2} \cdot \frac{\pi}{2}.$$

Analogamente substituindo em (15.54) n por $2k-1$ e multiplicando as equações que resultam para $k=1, 2, \dots, n$ obtem-se

$$\prod_{k=1}^n \frac{I_{2k+1}}{I_{2k-1}} = \prod_{k=1}^n \frac{(2k)^2}{2k(2k+1)} = \frac{2^{2n}(n!)^2}{(2n+1)!} = \frac{1}{2n+1} \cdot \frac{\pi}{2} \cdot \frac{1}{I_{2n}}.$$

O produto no primeiro membro significa-se para $I_{2n+1}/I_1 = I_{2n+1}$, pelo que obtemos

$$(15.56) \quad I_{2n}I_{2n+1} = \frac{\pi}{2(2n+1)}.$$

Visto que $I_{2n+1} = 2nI_{2n-1}/(2n+1)$, a equação (15.56) implica

$$I_{2n}I_{2n-1} = \frac{\pi}{4n}.$$

Utilizamos este resultado em (15.53), conjuntamente com as duas relações (15.55) e (15.56). Multiplicando por $\pi^2/4$ obtemos a desigualdade de Wallis (15.52).

15.23. Exercícios

1. Se f é um polinómio de grau ≤ 2 , mostrar que a fórmula de Euler (15.48) se reduz à fórmula trapezoidal.
2. A constante de Euler define-se pela fórmula limite

$$C = \lim_{n \rightarrow \infty} \left(\sum_{k=1}^n \frac{1}{k} - \log n \right).$$

(Ver Secção 10.17 do Volume I). Utilizar a fórmula de Euler para provar que

$$\sum_{k=1}^n \frac{1}{k} = \log n + C + \frac{1}{2n} - \frac{E(n)}{n^2},$$

onde $0 \leq E(n) \leq \frac{1}{8}$. Mostrar também que

$$C = 1 - \int_1^{\infty} \frac{t - [t]}{t^2} dt.$$

3. (a) Se $s > 0$, $s \neq 1$, usar a fórmula de Euler para provar que

$$\sum_{k=1}^n \frac{1}{k^s} = \frac{n^{1-s}}{1-s} + C(s) + s \int_n^\infty \frac{t - [t]}{t^{s+1}} dt,$$

sendo

$$C(s) = 1 + \frac{1}{s-1} - s \int_1^\infty \frac{t - [t]}{t^{s+1}} dt.$$

- (b) Se $s > 1$, demonstrar que $C(s) = \zeta(s)$, sendo ζ a função zeta de Riemann definida, para $s > 1$, pela série

$$\zeta(s) = \sum_{k=1}^{\infty} \frac{1}{k^s}.$$

Esta série $\zeta(s)$ diverge para $s \leq 1$. Todavia, uma vez que a fórmula para $C(s)$ da alínea (a) tem significado para $0 < s < 1$, pode usar-se para generalizar a definição de $\zeta(s)$ ao intervalo aberto $0 < s < 1$. Assim, para $s > 0$ e $s \neq 1$ temos

$$\zeta(s) = 1 + \frac{1}{s-1} - s \int_1^\infty \frac{t - [t]}{t^{s+1}} dt.$$

Isto é um *teorema* se $s > 1$, e uma *definição* se $0 < s < 1$.

Nos Exercícios 4 a 6, φ_2 é a função introduzida em 15.22.

4. (a) Empregar a fórmula de somação de Euler para demonstrar que

$$\sum_{k=1}^n \log^2 k = (n + \frac{1}{2}) \log^2 n - 2n \log n + 2n - 2 + 2 \int_1^n \varphi_2(x) \frac{\log x - 1}{x^2} dx.$$

- (b) Utilizar a alínea (a) para deduzir que para $n > e$ se tem

$$\sum_{k=1}^n \log^2 k = (n + \frac{1}{2}) \log^2 n - 2n \log n + 2n + A - E(n),$$

sendo A uma constante e $0 < E(n) < \frac{\log n}{4n}$.

5. (a) Utilizar a fórmula de somação de Euler para provar que

$$\sum_{k=1}^n \frac{\log k}{k} = \frac{1}{2} \log^2 n + \frac{1}{2} \frac{\log n}{n} - \int_1^n \frac{2 \log x - 3}{x^3} \varphi_2(x) dx.$$

- (b) Servir-se da alínea (a) para demonstrar que para $n > e^{3/2}$ se tem

$$\sum_{k=1}^n \frac{\log k}{k} = \frac{1}{2} \log^2 n + \frac{1}{2} \frac{\log n}{n} + A - E(n),$$

onde A é constante e $0 < E(n) < \frac{\log n}{8n^2}$.

6. (a) Se $n > 2$ utilizar a fórmula de Euler para provar que

$$\sum_{k=2}^n \frac{1}{k \log k} = \log(\log n) + \frac{1}{2n \log n} + \frac{1}{4 \log 2} - \log(\log 2) - \int_2^n \varphi_2(x) \frac{2 + 3 \log x + 2 \log^2 x}{(x \log x)^3} dx$$

- (b) Utilizar a alínea (a) para demonstrar que para $n > 2$ se tem

$$\sum_{k=2}^n \frac{1}{k \log k} = \log(\log n) + A + \frac{1}{2n \log n} - E(n),$$

sendo A uma constante e $0 < E(n) < \frac{1}{4n^2 \log n}$.

7. (a) Se $a > 0$ e $p > 0$, demonstrar, recorrendo à fórmula de Euler, que

$$\sum_{k=0}^{\infty} e^{-ak^p} = \frac{\Gamma\left(1 + \frac{1}{p}\right)}{a^{1/p}} + \frac{1}{2} - ap \int_0^{\infty} \varphi_1(x) x^{p-1} e^{-ax^p} dx,$$

onde Γ é a função gamma.

- (b) Utilizar a alínea (a) para demonstrar

$$\sum_{k=0}^{\infty} e^{-ak^p} = \frac{\Gamma\left(1 + \frac{1}{p}\right)}{a^{1/p}} + \theta, \quad \text{com } 0 < \theta < 1.$$

8. Demonstrar os limites seguintes com a ajuda da fórmula de Stirling e ou a desigualdade de Wallis

(a) $\lim_{n \rightarrow \infty} \frac{n}{(n!)^{1/n}} = e.$

(b) $\lim_{n \rightarrow \infty} \frac{(n!)^{3/2} 2^{2n}}{(2n)! \sqrt{n}} = \sqrt{\pi}.$

(c) $\lim_{n \rightarrow \infty} (-1)^n \binom{-\frac{1}{2}}{n} n = \frac{1}{\sqrt{\pi}}.$

9. Seja $I_n = \int_0^{\pi/2} \sin^n t \, dt$, com n um inteiro não negativo. Na Secção 15.22 mostrou-se que a sucessão $\{I_n\}$ satisfaz à fórmula

$$I_{n+2} = \frac{n+1}{n+2} I_n.$$

Seja $f(n) = \frac{1}{2} \sqrt{\pi} \Gamma\left(\frac{n+1}{2}\right) / \Gamma\left(\frac{n}{2} + 1\right)$, onde Γ é a função gama.

(a) Utilizar a equação funcional $\Gamma(s+1) = s \Gamma(s)$ para provarmos que

$$f(n+2) = \frac{n+1}{n+2} f(n).$$

(b) Utilizar a alínea (a) para demonstrar que

$$\int_0^{\pi/2} \sin^n t \, dt = \frac{\sqrt{\pi}}{2} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2} + 1\right)}.$$

BIBLIOGRAFIA

Na reduzida lista que a seguir se apresenta figuram alguns livros sugeridos para uma leitura suplementar dos princípios gerais de análise numérica. Todos estes livros contêm referências a trabalhos de índole mais especializada. Recomenda-se em especial a lista de tábuas apresentadas na referência número nove.

Livros

1. A. D. Booth, *Numerical Methods*, Academic Press, New York, 1958; 3.^a edição Plenum Press, New York, 1966.
2. P. J. Davis, *Interpolation and Approximation*, Blaisdell, Waltham, Mass., 1963.
3. D. R. Hartree, *Numerical Analysis*, Oxford Univ. Press (clarendon), Londres e New York, 1958.
4. F. B. Hildebrand, *Introduction to Numerical Analysis*, Mc Graw-Hill, New York, 1956.
5. A. S. Householder, *Principles of Numerical Analysis*, Mc Graw-Hill, New York, 1953.
6. W. E. Milne, *Numerical Calculus*, Princeton Univ. Press, Princeton, N. J., 1950.
7. J. B. Scarborough, *Numerical Mathematical Analysis*, Johns Hopkins Press, Baltimore, 6.^a edição, 1966.
8. J. Todd, *Introduction to the Constructive Theory of Functions*, Academic Press, New York, 1963.
9. J. Todd (ed.), *Survey of Numerical Analysis*, Mc Graw-Hill, New York, 1962.

Tábuas

10. L. J. Comrie (ed.), *Chamber's Six-figure Mathematical Tables*, W. and R. Chambers, Londres e Edimburg, 1949.
11. L. J. Comrie, *Interpolation and Allied Tables*, 2.^a reimpressão do *Nautical Almanac* para 1937, H. M. Stationery Office, Londres, 1948.
12. A. J. Fletcher, J. C. P. Miller, L. Rosenhead, *Index of Mathematical Tables*, Mc Graw-Hill, New York, 1946.
13. *Tables of Lagrangian Interpolation Coefficients*, Natl. Bur. Standards Columbia Press Series, Vol. 4, Columbia Univ. Press, New York, 1944.

SOLUÇÕES DOS EXERCÍCIOS

Capítulo 1

1.5. Exercícios (pg. 8)

- | | | | |
|--------|---------|---------|---------|
| 1. Sim | 8. Sim | 15. Sim | 22. Sim |
| 2. Sim | 9. Sim | 16. Sim | 23. Não |
| 3. Sim | 10. Sim | 17. Sim | 24. Sim |
| 4. Sim | 11. Não | 18. Sim | 25. Não |
| 5. Não | 12. Sim | 19. Sim | 26. Sim |
| 6. Sim | 13. Sim | 20. Sim | 27. Sim |
| 7. Sim | 14. Não | 21. Sim | 28. Sim |
31. (a) Não (b) Não (c) Não (d) Não

1.10. Exercícios (pg. 15)

- | | | | |
|-----------|-----------|--------------|--------------|
| 1. Sim; 2 | 5. Sim; 1 | 9. Sim; 1 | 13. Sim; n |
| 2. Sim; 2 | 6. Não | 10. Sim; 1 | 14. Sim; n |
| 3. Sim; 2 | 7. Não | 11. Sim; n | 15. Sim; n |
| 4. Sim; 2 | 8. Não | 12. Sim; n | 16. Sim; n |
17. Sim; $\dim = 1 + \frac{1}{2}n$ se n é par, $\frac{1}{2}(n+1)$ se n é ímpar
 18. Sim; $\dim = \frac{1}{2}n$ se n é par; $\frac{1}{2}(n+1)$ se n é ímpar
 19. Sim; $k+1$
 20. Não
 21. (a) $\dim = 3$ (b) $\dim = 3$ (c) $\dim = 2$ (d) $\dim = 2$
 23. (a) Se $a \neq 0$ e $b \neq 0$, o conjunto é independente; $\dim = 3$; se a ou b é zero, o conjunto é dependente; $\dim = 2$ (b) independente; $\dim = 2$ (c) se $a \neq 0$, independente; $\dim = 3$; se $a = 0$ dependente; $\dim = 2$ (d) independente; $\dim = 3$ (e) dependente; $\dim = 2$ (f) independente; $\dim = 2$ (g) independente; $\dim = 2$ (h) dependente; $\dim = 2$ (i) independente; $\dim = 2$ (j) independente; $\dim = 2$

1.13. Exercícios (pg. 23)

1. (a) Não (b) Não (c) Não (d) Não (e) Sim
8. (a) $\frac{1}{2}\sqrt{e^2 + 1}$ (b) $g(x) = b\left(x - \frac{e^2 + 1}{4}\right)$, b arbitrário
10. (b) $\frac{(n+1)(2n+1)}{6n}a + \frac{n+1}{2}b$ (c) $g(t) = a\left(t - \frac{2n+1}{3n}\right)$, a arbitrário
11. (c) 43 (d) $g(t) = a(1 - \frac{2}{3}t)$, a arbitrário
12. (a) Não (b) Não (c) Não (d) Não
13. (c) 1 (d) $e^2 - 1$
14. (c) $n!/2^{n+1}$

1.17. Exercícios (pg. 34)

1. (a) e (b) $\frac{1}{3}\sqrt{3}(1, 1, 1)$, $\frac{1}{6}\sqrt{6}(1, -2, 1)$
2. (a) $\frac{1}{2}\sqrt{2}(1, 1, 0, 0)$, $\frac{1}{6}\sqrt{6}(-1, 1, 2, 0)$, $\frac{1}{6}\sqrt{3}(1, -1, 1, 3)$
 (b) $\frac{1}{3}\sqrt{3}(1, 1, 0, 1)$, $\frac{1}{\sqrt{42}}(1, -2, 6, 1)$
6. $\frac{2}{3} - \frac{1}{2}\log^2 3$ 9. $\pi - 2 \sin x$
7. $e^2 - 1$ 10. $\frac{3}{4} - \frac{1}{4}x$
8. $\frac{1}{2}(e - e^{-1}) + \frac{3}{e}x$; $1 - 7e^{-2}$

Capítulo 2

2.4. Exercícios (pg. 39)

1. Linear; nulidade 0, ordem 2
2. Linear; nulidade 0, ordem 2
3. Linear; nulidade 1, ordem 1
4. Linear; nulidade 1, ordem 1
5. Não linear
6. Não linear
7. Não linear
8. Não linear
9. Linear; nulidade 0, ordem 2
10. Linear; nulidade 0, ordem 2
11. Linear; nulidade 0, ordem 2
12. Linear; nulidade 0, ordem 2
13. Não linear
14. Linear; nulidade 0, ordem 2
15. Não linear
16. Linear; nulidade 0, ordem 2
17. Linear; nulidade 1, ordem 2
18. Linear; nulidade 0, ordem 3
19. Não linear
20. Não linear
21. Não linear
22. Não linear
23. Linear; nulidade 1, ordem 2
24. Linear; nulidade 0, ordem $n+1$
25. Linear; nulidade 1, ordem infinita
26. Linear; nulidade infinita, ordem 2
27. Linear; nulidade 2, ordem infinita
28. $N(T)$ é o conjunto das sucessões constantes; $T(V)$ é o conjunto das sucessões com limite 0.
29. (d) $\{1, \cos x, \sin x\}$ é uma base para $T(V)$; $\dim T(V) = 3$ (e) $N(T) = S$ (f) Se $T(f) = cf$ sendo $c \neq 0$, então $c \in T(V)$ pelo que se tem $f(x) = c_1 + c_2 \cos x + c_3 \sin x$; se $c_1 = 0$, então $c = \pi$ e $f(x) = c_1 \cos x + c_2 \sin x$, onde c_1 e c_2 são arbitrários, mas não si-

multaneamente nulos; se $c_1 \neq 0$, então $c = 2\pi$ e $f(x) = c_1$, onde c_1 é arbitrário, mas não nulo.

2.8. Exercícios (pg. 48)

3. Sim; $x = v$, $y = u$
4. Sim; $x = u$, $y = -v$
5. Não;
6. Não
7. Não;
8. Sim; $x = \log u$, $y = \log v$
9. Não;
10. Sim $x = u - 1$, $y = v - 1$
11. Sim; $x = \frac{1}{2}(v + u)$, $y = \frac{1}{2}(v - u)$
12. Sim; $x = \frac{1}{3}(v + u)$, $y = \frac{1}{3}(2v - u)$
13. Sim; $x = w$, $y = v$, $z = u$
14. Não;
15. Sim; $x = u$, $y = \frac{1}{2}v$, $z = \frac{1}{3}w$
16. Sim; $x = u$, $y = v$, $z = w - u - v$
17. Sim; $x = u - 1$, $y = v - 1$, $z = w + 1$
18. Sim; $x = u - 1$, $y = v - 2$, $z = w - 3$
19. Sim; $x = u$, $y = v - u$, $z = w - v$
20. Sim; $x = \frac{1}{2}(u - v + w)$, $y = \frac{1}{2}(v - w + u)$, $z = \frac{1}{2}(w - u + v)$
25. $(S + T)^2 = S^2 + ST + TS + T^2$;
 $(S + T)^3 = S^3 + TS^2 + STS + S^2T + ST^2 + TST + T^2S + T^3$
26. (a) $(ST)(x, y, z) = (x + y + z, x + y, x)$; $(TS)(x, y, z) = (z, z + y, z + y + x)$;
 $(ST - TS)(x, y, z) = (x + y, x - z, -y - z)$; $S^2(x, y, z) = (x, y, z)$;
 $T^2(x, y, z) = (x, 2x + y, 3x + 2y + z)$;
 $(ST)^2(x, y, z) = (3x + 2y + z, 2x + 2y + z, x + y + z)$;
 $(TS)^2(x, y, z) = (x + y + z, x + 2y + 2z, x + 2y + 3z)$;
 $(ST - TS)^2 = (2x + y - z, x + 2y + z, -x + y + 2z)$
 (b) $S^{-1}(u, v, w) = (w, v, u)$; $T^{-1}(u, v, w) = (u, v - u, w - v)$;
 $(ST)^{-1}(u, v, w) = (w, v - w, u - v)$; $(TS)^{-1}(u, v, w) = (w - v, v - u, u)$
 (c) $(T - I)(x, y, z) = (0, x, x + y)$; $(T - I)^2(x, y, z) = (0, 0, x)$;
 $(T - I)^n(x, y, z) = (0, 0, 0)$ se $n \geq 3$
28. (a) $Dp(x) = 3 - 2x + 12x^2$; $Tp(x) = 3x - 2x^2 + 12x^3$; $(DT)p(x) = 3 - 4x + 36x^2$;
 $(TD)p(x) = -2x + 24x^2$; $(DT - TD)p(x) = 3 - 2x + 12x^2$;
 $(T^2D^2 - D^2T^2)p(x) = 8 - 192x$
 (b) $p(x) = ax$, a um escalar qualquer
 (c) $p(x) = ax^2 + b$, a e b escalares quaisquer.
31. (a) $Rp(x) = 2$; $Sp(x) = 3 - x + x^2$; $Tp(x) = 2x + 3x^2 - x^3 + x^4$;
 $(ST)p(x) = 2 + 3x - x^2 + x^3$; $(TS)p(x) = 3x - x^2 + x^3$; $(TS)^2p(x) = 3x - x^2 + x^3$;
 $(T^2S^2)p(x) = -x^2 + x^3$; $(S^2T^2)p(x) = 2 + 3x - x^2 + x^3$; $(TRS)p(x) = 3x$;
 $(RST)p(x) = 2$ (b) $N(R) = \{p \mid p(0) = 0\}$; $R(V) = \{p \mid p \text{ é constante}\}$; $N(S) = \{p \mid p \text{ é constante}\}$; $S(V) = V$; $N(T) = \{0\}$; $T(V) = \{p \mid p(0) = 0\}$ (c) $T^{-1} = S$
 (d) $(TS)^n = I - R$; $S^n T^n = I$
32. T não é biunívoca em V porque aplica toda a sucessão constante na mesma sucessão.

2.12. Exercícios (pg. 56)

- 1 (a) A matriz identidade $I = (\delta_{jk})$, sendo $\delta_{jk} = 1$ se $j = k$ e $\delta_{jk} = 0$ se $j \neq k$.
 (b) A matriz zero $O = (a_{jk})$, em que cada elemento $a_{jk} = 0$.
 (c) A matriz $(c\delta_{jk})$, sendo (δ_{jk}) a matriz identidade da alínea (a).

$$2. \quad (a) \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad (b) \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (c) \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$3. \quad (a) -5i + 7j, \quad 9i - 12j$$

$$(b) \begin{bmatrix} 1 & 2 \\ 1 & -1 \end{bmatrix}, \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix} \quad (c) \begin{bmatrix} -\frac{7}{4} & -\frac{1}{4} \\ \frac{1}{4} & \frac{7}{4} \end{bmatrix}, \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$$

$$4. \quad \begin{bmatrix} -2 & 0 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$$

$$5. \quad (a) 3i + 4j + 4k; \text{ nulidade } 0, \text{ ordem } 3 \quad (b) \begin{bmatrix} -1 & -1 & 2 \\ 1 & -3 & 3 \\ -1 & -5 & 5 \end{bmatrix}$$

$$6. \quad \begin{bmatrix} 2 & 0 & -2 \\ 1 & -1 & 1 \\ 2 & 1 & 0 \end{bmatrix}$$

$$7. \quad (a) T(4i - j + k) = (0, -2); \text{ nulidade } 1, \text{ ordem } 2 \quad (b) \begin{bmatrix} 0 & 1 & 1 \\ 0 & 1 & -1 \end{bmatrix}$$

$$(c) \begin{bmatrix} 0 & 1 & 3 \\ 0 & 0 & -2 \end{bmatrix} \quad (d) e_1 = j, \quad e_2 = k, \quad e_3 = i, \quad w_1 = (1, 1), \quad w_2 = (1, -1)$$

$$8. \quad (a) (5, 0, -1); \text{ nulidade } 0, \text{ ordem } 2 \quad (b) \begin{bmatrix} 1 & -1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix}$$

$$(c) e_1 = i, \quad e_2 = i + j, \quad w_1 = (1, 0, 1), \quad w_2 = (0, 0, 2), \quad w_3 = (0, 1, 0)$$

$$9. \quad (a) (-1, -3, -1); \text{ nulidade } 0, \text{ ordem } 2 \quad (b) \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$$

$$(c) e_1 = i, \quad e_2 = j - i, \quad w_1 = (1, 0, 1), \quad w_2 = (0, 1, 0), \quad w_3 = (0, 0, 1)$$

$$10. \quad (a) e_1 - e_2; \text{ nulidade } 0, \text{ ordem } 2 \quad (b) \begin{bmatrix} 1 & 2 \\ 5 & 4 \end{bmatrix} \quad (c) a = 5, \quad b = 4$$

$$11. \quad \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$$

$$12. \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$13. \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$14. \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$$

$$15. \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$$

$$16. \begin{bmatrix} 0 & -1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} -1 & 0 & 0 & -2 \\ 0 & -1 & 2 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

$$17. \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & -2 \\ 2 & 0 \end{bmatrix}$$

$$18. \begin{bmatrix} 2 & -3 \\ 3 & 2 \end{bmatrix}, \begin{bmatrix} -5 & -12 \\ 12 & -5 \end{bmatrix}$$

$$19. \text{(a)} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} \quad \text{(b)} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 9 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{(c)} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 6 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\text{(d)} \begin{bmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & -3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{(e)} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 9 \end{bmatrix} \quad \text{(f)} \begin{bmatrix} 0 & 0 & -8 & 0 \\ 0 & 0 & 0 & -48 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

20. Escolher $(x^3, x^2, x, 1)$ como uma base de V , e (x^2, x) como uma base de W . Então a matriz de TD é

$$\begin{bmatrix} 6 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \end{bmatrix}$$

2.16. Exercícios (pg. 64)

$$1. \quad B + C = \begin{bmatrix} 3 & 4 \\ 0 & 2 \\ 6 & -5 \end{bmatrix}, \quad AB = \begin{bmatrix} 15 & -14 \\ -15 & 14 \end{bmatrix}, \quad BA = \begin{bmatrix} -1 & 4 & -2 \\ -4 & 16 & -8 \\ 7 & -28 & 14 \end{bmatrix},$$

$$AC = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad CA = \begin{bmatrix} 0 & 0 & 0 \\ 2 & -8 & 4 \\ 4 & -16 & 8 \end{bmatrix}, \quad A(2B - 3C) = \begin{bmatrix} 30 & -28 \\ -30 & 28 \end{bmatrix}$$

$$2. \quad (a) \begin{bmatrix} a & b \\ 0 & 0 \end{bmatrix}, \quad a \text{ e } b \text{ arbitrários} \quad (b) \begin{bmatrix} -2a & a \\ -2b & b \end{bmatrix}, \quad a \text{ e } b \text{ arbitrários.}$$

$$3. \quad (a) \quad a = 9, \quad b = 6, \quad c = 1, \quad d = 5 \quad (b) \quad a = 1, \quad b = 6, \quad c = 0, \quad d = -2$$

$$4. \quad (a) \begin{bmatrix} -9 & -2 & -10 \\ 6 & 14 & 8 \\ -7 & 5 & -5 \end{bmatrix} \quad (b) \begin{bmatrix} -3 & 5 & -4 \\ 0 & 3 & 24 \\ 12 & -27 & 0 \end{bmatrix}$$

$$6. \quad A^n = \begin{bmatrix} 1 & n \\ 0 & 1 \end{bmatrix}$$

$$7. \quad A^n = \begin{bmatrix} \cos n\theta & -\operatorname{sen} n\theta \\ \operatorname{sen} n\theta & \cos n\theta \end{bmatrix}$$

$$8. \quad A^n = \begin{bmatrix} 1 & n & \frac{n(n+1)}{2} \\ 0 & 1 & n \\ 0 & 0 & 1 \end{bmatrix}$$

$$9. \quad \begin{bmatrix} 1 & 0 \\ -100 & 1 \end{bmatrix}$$

$$10. \quad \begin{bmatrix} a & b \\ c & -a \end{bmatrix}, \quad \text{sendo } b \text{ e } c \text{ arbitrários e } a \text{ uma solução qualquer da equação } a^2 = -bc.$$

$$11. \quad (b) \quad \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}, \quad \text{onde } a \text{ é arbitrário}$$

$$12. \quad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \text{e} \quad \begin{bmatrix} a & b \\ c & -a \end{bmatrix}, \quad \text{sendo } b \text{ e } c \text{ arbitrários e } a \text{ uma solução qualquer da equação } a^2 = 1 - bc.$$

$$13. C = \begin{bmatrix} \frac{15}{2} & \frac{13}{2} \\ 8 & 7 \end{bmatrix}, \quad D = \begin{bmatrix} \frac{33}{4} & \frac{19}{4} \\ \frac{43}{4} & \frac{25}{4} \end{bmatrix}$$

14. (b) $(A+B)^2 = A^2 + AB + BA + B^2$; $(A+B)(A-B) = A^2 + BA - AB - B^2$
 (c) Para aqueles que comutam.

2.20. Exercícios (pg. 76)

1. $(x, y, z) = (\frac{8}{5}, -\frac{7}{5}, \frac{8}{5})$
2. Nenhuma solução.
3. $(x, y, z) = (1, -1, 0) + t(-3, 4, 1)$
4. $(x, y, z) = (1, -1, 0) + t(-3, 4, 1)$
5. $(x, y, z, u) = (1, 1, 0, 0) + t(1, 14, 5, 0)$
6. $(x, y, z, u) = (1, 8, 0, -4) + t(2, 7, 3, 0)$
7. $(x, y, z, u, v) = t_1(-1, 1, 0, 0, 0) + t_2(-1, 0, 3, -3, 1)$
8. $(x, y, z, u) = (1, 1, 1, -1) + t_1(-1, 3, 7, 0) + t_2(4, 9, 0, 7)$
9. $(x, y, z) = (\frac{4}{3}, \frac{2}{3}, 0) + t(5, 1, -3)$
10. (a) $(x, y, z, u) = (1, 6, 3, 0) + t_1(4, 11, 7, 0) + t_2(0, 0, 0, 1)$
 (b) $(x, y, z, u) = (\frac{3}{11}, 4, \frac{19}{11}, 0) + t(4, -11, 7, 22)$

$$12. \begin{bmatrix} -1 & 2 & 1 \\ 5 & -8 & -6 \\ -3 & 5 & 4 \end{bmatrix}$$

$$14. \begin{bmatrix} 14 & 8 & 3 \\ 8 & 5 & 2 \\ 3 & 2 & 1 \end{bmatrix}$$

$$13. \begin{bmatrix} -\frac{5}{8} & \frac{2}{3} & \frac{4}{3} \\ -1 & 0 & 1 \\ \frac{7}{3} & -\frac{1}{3} & -\frac{5}{3} \end{bmatrix}$$

$$15. \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$16. \begin{bmatrix} 0 & \frac{1}{2} & 0 & -1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \\ -3 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \\ 9 & 0 & -3 & 0 & 1 & 0 \end{bmatrix}$$

2.21. Exercícios variados sobre matrizes (pg. 77)

$$3. P = \begin{bmatrix} 2 & 1 \\ 5 & -1 \end{bmatrix}$$

4. $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, e $\begin{bmatrix} a & b \\ c & 1-a \end{bmatrix}$, onde b e c são arbitrários e a é uma solução qualquer da equação do segundo grau $a^2 - a + bc = 0$.

$$10. \quad (a) \quad \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} -1 & -1 \\ 1 & -1 \end{bmatrix}, \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix},$$

$$\begin{bmatrix} 1 & -1 \\ -1 & -1 \end{bmatrix}, \begin{bmatrix} -1 & 1 \\ -1 & -1 \end{bmatrix}.$$

Capítulo 3

3.6. Exercícios (pg. 89)

1. (a) 6 (b) 76 (c) $a^3 - 4a$
 2. (a) 1 (b) 1 (c) 1
 3. (b) $(b-a)(c-a)(c-b)(a+b+c)$ and $(b-a)(c-a)(c-b)(ab+ac+bc)$
 4. (a) 8 (b) $(b-a)(c-a)(d-a)(c-b)(d-b)(d-c)$
 (c) $(b-a)(c-a)(d-a)(c-b)(d-b)(d-c)(a+b+c+d)$
 (d) $a(a^2-4)(a^2-16)$ (e) -160

$$7. \quad F' = \begin{vmatrix} f'_1 & f'_2 & f'_3 \\ g'_1 & g'_2 & g'_3 \\ h'_1 & h'_2 & h'_3 \end{vmatrix} + \begin{vmatrix} f_1 & f_2 & f_3 \\ g'_1 & g'_2 & g'_3 \\ h_1 & h_2 & h_3 \end{vmatrix} + \begin{vmatrix} f_1 & f_2 & f_3 \\ g_1 & g_2 & g_3 \\ h'_1 & h'_2 & h'_3 \end{vmatrix}$$

$$8. \quad (b) \quad \text{Se } F = \begin{vmatrix} f_1 & f_2 & f_3 \\ f'_1 & f'_2 & f'_3 \\ f''_1 & f''_2 & f''_3 \end{vmatrix} \text{ então } F' = \begin{vmatrix} f_1 & f_2 & f_3 \\ f'_1 & f'_2 & f'_3 \\ f'''_1 & f'''_2 & f'''_3 \end{vmatrix}$$

$$10 \quad \det A = 16, \quad \det(A^{-1}) = \frac{1}{16}, \quad A^{-1} = \begin{bmatrix} \frac{1}{2} & -\frac{3}{4} & \frac{1}{8} & \frac{1}{16} \\ 0 & \frac{1}{2} & -\frac{3}{4} & \frac{1}{8} \\ 0 & 0 & \frac{1}{2} & -\frac{3}{4} \\ 0 & 0 & 0 & \frac{1}{2} \end{bmatrix}$$

3.11. Exercícios (pg. 95)

6. $\det A = (\det B)(\det D)$
 7. (a) Independentes (b) Independentes (c) Dependentes

3.17. Exercícios (pg. 106)

$$1. \quad (a) \quad \begin{bmatrix} 4 & -3 \\ -2 & 1 \end{bmatrix} \quad (b) \quad \begin{bmatrix} 2 & -1 & 1 \\ -6 & 3 & 5 \\ -4 & -2 & 2 \end{bmatrix} \quad (c) \quad \begin{bmatrix} 109 & 113 & -41 & -13 \\ -40 & -92 & 74 & 16 \\ -41 & -79 & 7 & 47 \\ -50 & 38 & 16 & 20 \end{bmatrix}$$

$$2. (a) -\frac{1}{2} \begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix} \quad (b) \frac{1}{8} \begin{bmatrix} 2 & -6 & -4 \\ -1 & 3 & -2 \\ 1 & 5 & 2 \end{bmatrix} \quad (c) \frac{1}{306} \begin{bmatrix} 109 & -40 & -41 & -50 \\ 113 & -92 & -79 & 38 \\ -41 & 74 & 7 & 16 \\ -13 & 16 & 47 & 20 \end{bmatrix}$$

$$3. (a) \lambda = 2, \lambda = -3 \quad (b) \lambda = 0, \lambda = \pm 3 \quad (c) \lambda = 3, \lambda = \pm i$$

$$5. (a) x = 0, y = 1, z = 2 \quad (b) x = 1, y = 1, z = -1$$

$$6. (b) \det \begin{bmatrix} x - x_1 & y - y_1 & z - z_1 \\ x_2 - x_1 & y_2 - y_1 & z_2 - z_1 \\ x_3 - x_1 & y_3 - y_1 & z_3 - z_1 \end{bmatrix} = 0; \det \begin{bmatrix} x & y & z & 1 \\ x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \end{bmatrix} = 0$$

$$(c) \det \begin{bmatrix} (x - x_1)^2 + (y - y_1)^2 & (x - x_1) & (y - y_1) \\ (x_2 - x_1)^2 + (y_2 - y_1)^2 & (x_2 - x_1) & (y_2 - y_1) \\ (x_3 - x_1)^2 + (y_3 - y_1)^2 & (x_3 - x_1) & (y_3 - y_1) \end{bmatrix} = 0;$$

$$\det \begin{bmatrix} x^2 + y^2 & x & y & 1 \\ x_1^2 + y_1^2 & x_1 & y_1 & 1 \\ x_2^2 + y_2^2 & x_2 & y_2 & 1 \\ x_3^2 + y_3^2 & x_3 & y_3 & 1 \end{bmatrix} = 0$$

Capítulo 4

4.4. Exercícios (pg. 115)

5. Funções próprias: $f(t) = Ct^\lambda$, com $C \neq 0$
6. O polinômio constante não nulo
7. Funções próprias: $f(t) = Ce^{t/\lambda}$, com $C \neq 0$
8. Funções próprias: $f(t) = Ce^{1/2t/\lambda}$, com $C \neq 0$
10. Vectors próprios correspondentes a $\lambda = 0$ são todos sucessões constantes com limite $a \neq 0$. Vectors próprios correspondentes a $\lambda = -1$ são sucessões não constantes com limite $a = 0$.

4.8. Exercícios (pg. 121)

	Valor próprio	Vector próprio	$\dim E(\lambda)$
1.	(a) 1, 1	$(a, b) \neq (0, 0)$	2
	(b) 1, 1	$t(1, 0), t \neq 0$	1
	(c) 1, 1	$t(0, 1), t \neq 0$	1
	(d) 2	$t(1, 1), t \neq 0$	1
	0	$t(1, -1), t \neq 0$	1
2.	$1 + \sqrt{ab}$	$t(\sqrt{a}, \sqrt{b}), t \neq 0$	1
	$1 - \sqrt{ab}$	$t(\sqrt{a}, -\sqrt{b}), t \neq 0$	1

3. Se o campo de escalares é o conjunto dos números reais \mathbf{R} , então valores próprios reais existem unicamente quando $\sin \theta = 0$, caso em que existirão dois valores próprios iguais, $\lambda_1 = \lambda_2 = \cos \theta$, com $\cos \theta = 1$ ou -1 . Neste caso todo o vector não nulo é um vector próprio, pelo que $\dim E(\lambda_1) = \dim E(\lambda_2) = 2$.

Se o campo de escalares é o conjunto dos números complexos \mathbf{C} , então os valores próprios são $\lambda_1 = \cos \theta + i \sin \theta$, $\lambda_2 = \cos \theta - i \sin \theta$. Se $\sin \theta = 0$ aqueles são reais e iguais. Se $\sin \theta \neq 0$ são complexos conjugados; os vectores próprios correspondentes a λ_1 são $t(i, 1)$, $t \neq 0$; os correspondentes a λ_2 são $t(1, i)$, $t \neq 0$; $\dim E(\lambda_1) = \dim E(\lambda_2) = 1$.

4. $\begin{bmatrix} a & b \\ c & -a \end{bmatrix}$, onde b e c são arbitrários e a uma das soluções da equação $a^2 = 1 - bc$.
5. Seja $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, e $\Delta = (a - d)^2 + 4bc$. Os valores próprios são reais e distintos se $\Delta > 0$, reais e iguais se $\Delta = 0$, conjugados complexos se $\Delta < 0$.
6. $a = b = c = d = e = f = 1$.

	Valor próprio	Vector próprio	$\dim E(\lambda)$
7. (a)	1, 1, 1	$t(0, 0, 1)$, $t \neq 0$	1
(b)	1	$t(1, -1, 0)$, $t \neq 0$	1
	2	$t(3, 3, -1)$, $t \neq 0$	1
	21	$t(1, 1, 6)$, $t \neq 0$	1
(c)	1	$t(3, -1, 3)$, $t \neq 0$	1
	2, 2	$t(2, 2, -1)$, $t \neq 0$	1

8. 1, 1, -1, -1 para cada matriz.

4.10. Exercícios (pg. 127)

2. (a) Valores próprios 1, 3; $C = \begin{bmatrix} -2c & 0 \\ c & d \end{bmatrix}$, onde $cd \neq 0$
- (b) Valores próprios 6, -1; $C = \begin{bmatrix} 2a & b \\ 5a & -b \end{bmatrix}$, onde $ab \neq 0$
- (c) Valores próprios 3, 3; se existir uma matriz não singular C então $C^{-1}AC = 3I$, pelo que $AC = 3C$, $A = 3I$
- (d) Valores próprios 1, 1; se existir uma matriz não singular C então $C^{-1}AC = I$, pelo que $AC = C$, $A = I$.
3. $C = A^{-1}B$.
4. (a) Valores próprios 1, 1, -1; vectores próprios (1, 0, 1), (0, 1, 0), (1, 0, -1);

$$C = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & -1 \end{bmatrix}$$

- (b) Valores próprios 2, 2, 1; vectores próprios (1, 0, -1), (0, 1, -1), (1, -1, 1);

$$C = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix}$$

5. (a) Valores próprios 2, 2; vectores próprios $t(1, 0)$, $t \neq 0$. Se $C = \begin{bmatrix} a & b \\ -b & 0 \end{bmatrix}$, $b \neq 0$, então

$$C^{-1}AC = \begin{bmatrix} 2 & 0 \\ 1 & 2 \end{bmatrix}$$

- (b) Valores próprios 3, 3; vectores próprios $t(1, 1)$, $t \neq 0$. Se $C = \begin{bmatrix} a & b \\ a+b & b \end{bmatrix}$, $b \neq 0$, então

$$C^{-1}AC = \begin{bmatrix} 3 & 0 \\ 1 & 3 \end{bmatrix}$$

6. Valores próprios 1, 1, 1; vectores próprios $t(1, -1, -1)$, $t \neq 0$.

Capítulo 5

5.5. Exercícios (pg. 134)

3. (b) T^n é hermitica se n é par; se n é ímpar é hemi-hermitica.
 7. (a) Simétrica (b) Nem uma coisa nem outra (c) Simétrica (d) Simétrica
 9. (d) $Q(x + ty) = Q(x) + tQ(y) + i(T(x), y) + t(T(y), x)$

5.11. Exercícios (pg. 140)

1. (a) Simétrica e hermitica
 (b) Nenhum dos quatro tipos
 (c) Hemi-simétrica
 (d) Hemi-simétrica e hemi-hermitica
 4. (b) $\begin{bmatrix} \cos \theta & \sin \theta \\ \sin \theta & -\cos \theta \end{bmatrix}$
 5. Valores próprios $\lambda_1 = 0$, $\lambda_2 = 25$; vectores próprios ortonormados.

$$u_1 = \frac{1}{5}(4, -3), u_2 = \frac{1}{5}(3, 4).$$

$$C = \frac{1}{5} \begin{bmatrix} 4 & 3 \\ -3 & 4 \end{bmatrix}$$

6. Valores próprios $\lambda_1 = 2i$, $\lambda_2 = -2i$; vectores próprios ortonormados

$$u_1 = \frac{1}{\sqrt{2}}(1, -i), u_2 = \frac{1}{\sqrt{2}}(1, i). \quad C = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix}$$

7. Valores próprios $\lambda_1 = 1$, $\lambda_2 = 3$, $\lambda_3 = -4$; Vectores próprios ortonormados.

$$u_1 = \frac{1}{\sqrt{10}}(1, 0, 3), \quad u_2 = \frac{1}{\sqrt{14}}(3, 2, -1), \quad u_3 = \frac{1}{\sqrt{35}}(3, -5, -1).$$

$$C = \begin{bmatrix} \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{14}} & \frac{3}{\sqrt{35}} \\ 0 & \frac{2}{\sqrt{14}} & \frac{-5}{\sqrt{35}} \\ \frac{3}{\sqrt{10}} & \frac{-1}{\sqrt{14}} & \frac{-1}{\sqrt{35}} \end{bmatrix}$$

8. Valores próprios $\lambda_1 = 1$, $\lambda_2 = 6$, $\lambda_3 = -4$; vectores próprios ortonormados

$$u_1 = \frac{1}{5}(0, 4, -3), \quad u_2 = \frac{1}{\sqrt{50}}(5, 3, 4), \quad u_3 = \frac{1}{\sqrt{50}}(5, -3, -4).$$

$$C = \frac{1}{\sqrt{50}} \begin{bmatrix} 0 & 5 & 5 \\ 4\sqrt{2} & 3 & -3 \\ -3\sqrt{2} & 4 & -4 \end{bmatrix}$$

9. (a), (b), (c) são unitários; (b) e (c) são ortogonais.

11. (a) Valores próprios $\lambda_1 = ia$, $\lambda_2 = -ia$; vectores próprios ortonormados

$$u_1 = \frac{1}{\sqrt{2}}(1, i), \quad u_2 = \frac{1}{\sqrt{2}}(1, -i). \quad (b) \quad C = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}$$

5.15. Exercícios (pg. 710)

$$1. (a) \quad A = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} \quad (b) \quad \lambda_1 = 0, \quad \lambda_2 = 5 \quad (c) \quad u_1 = \frac{1}{\sqrt{5}}(1, -2), \quad u_2 = \frac{1}{\sqrt{5}}(2, 1)$$

$$(d) \quad C = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}$$

$$2. (a) \quad A = \begin{bmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{bmatrix} \quad (b) \quad \lambda_1 = \frac{1}{2}, \quad \lambda_2 = -\frac{1}{2}$$

$$(c) \quad u_1 = \frac{1}{\sqrt{2}}(1, 1), \quad u_2 = \frac{1}{\sqrt{2}}(1, -1) \quad (d) \quad C = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$3. (a) \quad A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad (b) \quad \lambda_1 = \sqrt{2}, \quad \lambda_2 = -\sqrt{2}$$

$$(c) \quad u_1 = t(1 + \sqrt{2}, 1), \quad u_2 = t(-1, 1 + \sqrt{2}), \quad \text{onde } t = 1/\sqrt{4 + 2\sqrt{2}}$$

$$(d) \quad C = t \begin{bmatrix} 1 + \sqrt{2} & -1 \\ 1 & 1 + \sqrt{2} \end{bmatrix}, \quad \text{onde } t = 1/\sqrt{4 + 2\sqrt{2}}$$

$$4. (a) \quad A = \begin{bmatrix} 34 & -12 \\ -12 & 41 \end{bmatrix} \quad (b) \quad \lambda_1 = 50, \quad \lambda_2 = 25$$

$$(c) \quad u_1 = \frac{1}{5}(3, -4), \quad u_2 = \frac{1}{5}(4, 3) \quad (d) \quad C = \frac{1}{5} \begin{bmatrix} 3 & 4 \\ -4 & 3 \end{bmatrix}$$

$$5. (a) \quad A = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \quad (b) \quad \lambda_1 = 0, \quad \lambda_2 = \frac{3}{2}, \quad \lambda_3 = -\frac{1}{2}$$

$$(c) \quad u_1 = \frac{1}{\sqrt{3}}(1, -1, -1), \quad u_2 = \frac{1}{\sqrt{6}}(2, 1, 1), \quad u_3 = \frac{1}{\sqrt{2}}(0, 1, -1)$$

$$(c) \quad C = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{2} & 2 & 0 \\ -\sqrt{2} & 1 & \sqrt{3} \\ -\sqrt{2} & 1 & -\sqrt{3} \end{bmatrix}$$

$$6. (a) \quad A = \begin{bmatrix} 2 & 0 & 2 \\ 0 & 1 & 0 \\ 2 & 0 & -1 \end{bmatrix} \quad (b) \quad \lambda_1 = 1, \quad \lambda_2 = 3, \quad \lambda_3 = -2$$

$$(c) \quad u_1 = (0, 1, 0), \quad u_2 = \frac{1}{\sqrt{5}}(2, 0, 1), \quad u_3 = \frac{1}{\sqrt{5}}(1, 0, -2)$$

$$(d) \quad C = \frac{1}{\sqrt{5}} \begin{bmatrix} 0 & 2 & 1 \\ \sqrt{5} & 0 & 0 \\ 0 & 1 & -2 \end{bmatrix}$$

$$7. (a) \quad A = \begin{bmatrix} 3 & 2 & 4 \\ 2 & 0 & 2 \\ 4 & 2 & 3 \end{bmatrix} \quad (b) \quad \lambda_1 = \lambda_2 = -1, \quad \lambda_3 = 8$$

$$(c) \quad u_1 = \frac{1}{\sqrt{2}}(1, 0, -1), \quad u_2 = \frac{1}{3\sqrt{2}}(-1, 4, -1), \quad u_3 = \frac{1}{3}(2, 1, 2)$$

$$(d) \quad C = \frac{1}{3\sqrt{2}} \begin{bmatrix} 3 & -1 & 2\sqrt{2} \\ 0 & 4 & \sqrt{2} \\ -3 & -1 & 2\sqrt{2} \end{bmatrix}$$

8. Elipse; centro em (0, 0)
9. Hipérbole; centro em $(-\frac{1}{2}, -\frac{1}{2})$
10. Parábola; centro em $(\frac{1}{16}, -\frac{1}{16})$
11. Elipse; centro em (0, 0)
12. Elipse; centro em (6, -4)
13. Parábola; vértice em $(\frac{2}{23}, \frac{11}{23})$
14. Elipse; centro em (0, 0)
15. Parábola; vértice em $(\frac{1}{2}, \frac{1}{2})$
16. Elipse; centro em $(-1, \frac{1}{2})$
17. Hipérbole; centro em (0, 0)
18. Hipérbole; centro em $(-1, 2)$
19. - 14

5.20. Exercícios (pg. 158)

8. $a = \pm \frac{1}{3}\sqrt{3}$
13. (a), (b) e (c)

Capítulo 6

6.3. Exercícios (pg. 164)

1. $y = e^{3x} - e^{2x}$
2. $y = \frac{2}{3}x^2 + \frac{1}{3}x^5$
3. $y = 4 \cos x - 2 \cos^2 x$
4. Quatro vezes a quantidade inicial.
5. $f(x) = Cx^n$, or $f(x) = Cx^{1/n}$
6. (b) $y = e^{4x} - e^{-x^{3/3}}$
7. $y = c_1 e^{2x} + c_2 e^{-2x}$
8. $y = c_1 \cos 2x + c_2 \sin 2x$
9. $y = e^x(c_1 \cos 2x + c_2 \sin 2x)$
10. $y = e^{-x}(c_1 + c_2 x)$
11. $k = n^2 \pi^2$; $f_k(x) = C \sin n\pi x$ ($n = 1, 2, 3, \dots$)
13. (a) $y'' - y = 0$
(b) $y'' - 4y' + 4y = 0$
(c) $y'' + y' + \frac{5}{4}y = 0$
(d) $y'' + 4y = 0$
(e) $y'' - y = 0$
14. $y = \frac{1}{3}\sqrt{6}$, $y'' = -12y = -4\sqrt{6}$

6.9. Exercícios (pg. 175)

1. $y = c_1 + c_2 e^{-x} + c_3 e^{3x}$
2. $y = c_1 + c_2 e^x + c_3 e^{-x}$
3. $y = c_1 + (c_2 + c_3 x)e^{-2x}$
4. $y = (c_1 + c_2 x + c_3 x^2)e^{-x}$
5. $y = (c_1 + c_2 x + c_3 x^2 + c_4 x^3)e^{-x}$
6. $y = c_1 e^{2x} + c_2 e^{-2x} + c_3 \cos 2x + c_4 \sin 2x$
7. $y = e^{\sqrt{2}x}(c_1 \cos \sqrt{2}x + c_2 \sin \sqrt{2}x) + e^{-\sqrt{2}x}(c_3 \cos \sqrt{2}x + c_4 \sin \sqrt{2}x)$
8. $y = c_1 e^x + e^{-x/2}(c_2 \cos \frac{1}{2}\sqrt{3}x + c_3 \sin \frac{1}{2}\sqrt{3}x)$
9. $y = e^{-x}[(c_1 + c_2 x) \cos x + (c_3 + c_4 x) \sin x]$
10. $y = (c_1 + c_2 x) \cos x + (c_3 + c_4 x) \sin x$
11. $y = c_1 + c_2 x + (c_3 + c_4 x) \cos \sqrt{2}x + (c_5 + c_6 x) \sin \sqrt{2}x$
12. $y = c_1 + c_2 x + (c_3 + c_4 x) \cos 2x + (c_5 + c_6 x) \sin 2x$
13. $f(x) = \frac{1}{2m^2} (e^{mx} - \cos mx - \sin mx)$
15. (a) $y^{(4)} - 5y'' + 4y = 0$
(b) $y''' + 6y'' + 12y' + 8y = 0$
(c) $y^{(4)} - 2y''' + y'' = 0$
(d) $y^{(4)} - 2y''' + y'' = 0$
(e) $y^{(5)} - 2y^{(4)} + y''' = 0$
(f) $y^{(4)} + 8y''' + 33y'' + 68y' + 52y = 0$
(g) $y^{(4)} - 2y''' + y'' = 0$
(h) $y^{(6)} + 4y'' = 0$

6.15. Exercícios (pg. 188)

1. $y_1 = -2x - x^2 - \frac{1}{3}x^3$
2. $y_1 = \frac{1}{4}xe^{2x}$
3. $y_1 = (x - \frac{4}{3})e^x$
4. $y_1 = \frac{1}{3}\sin x$
5. $y_1 = \frac{1}{2}x^2e^x + e^{2x}$
6. $y_1 = \frac{1}{2}xe^x$
7. $y_1 = x \cosh x$
8. $y_1 = \frac{1}{24}x^4e^{-x}$
9. $50y_1 = (11 - 5x)e^x \sin 2x + (2 - 10x)e^x \cos 2x$
10. $y_1 = -(\frac{5}{8}x + \frac{3}{8}x^2 + \frac{1}{12}x^3)e^{-x}$

$$12. y_1 = \frac{x^m e^{\alpha x}}{p_A^{(m)}(\alpha)}$$

$$15. (b) 2D \quad (c) 3D^2 \quad (d) nD^{n-1}$$

$$16. y = Ae^x + Be^{-x} + \frac{1}{2}e^x \int \frac{e^{-x}}{x} dx - \frac{1}{2}e^{-x} \int \frac{e^x}{x} dx$$

$$17. y = (A + \frac{1}{2}x)\sin 2x + (B + \frac{1}{4}\log |\cos 2x|)\cos 2x$$

$$18. y = Ae^x + Be^{-x} + \frac{1}{2}\sec x$$

$$19. y = (A + Bx)e^x + e^{e^x} - xe^x \int e^{e^x} dx + e^x \int xe^{e^x} dx$$

$$20. y = -\frac{1}{8}\log|x| + \frac{1}{3}e^x \int \frac{e^{-x}}{x} dx - \frac{1}{4}e^{2x} \int \frac{e^{-2x}}{x} dx \\ + \frac{1}{24}e^{4x} \int \frac{e^{-4x}}{x} dx + Ae^x + Be^{2x} + Ce^{4x}$$

6.16. Exercícios variados sobre equações diferenciais lineares (pg. 189)

1. $u(x) = 6(e^{4x} - e^{-x})/5$; $v(x) = e^x - e^{-5x}$
2. $u(x) = \frac{5}{2}e^{2x-\pi} \sin 5x$; $v(x) = \frac{5}{8}e^{-2x-\pi} \sin 3x$
3. $u(x) = e^{-x^2}$; $Q(x) = 4x^2 + 2$
5. $y = (A + Bx^3)e^x + (x^2 - 2x + 2)e^{2x}$
6. $y = Ae^{4x} \int e^{-4x-x^2/3} dx + Be^{4x}$
7. $y = Ax^{1/2} + Bx^{-1/2}$
8. $y = Ae^x + Bx^2e^{-x} - x$
9. $y = A(x^2 - 2) + B/x$
10. $y = x^{-2}[A + B(x-1)^3 + \frac{1}{9}x^3 + \frac{2}{3}x^2 - \frac{7}{6}x + \frac{1}{2} - (x-1)^3 \log|x-1|]$
11. $a = 1, -1$; $y = [Ae^{g(x)} + Be^{-g(x)}]/x$

6.21. Exercícios (pg. 200)

2. $f(x) = u_1(x)$ ($\alpha = 1$)
3. (a) $A = (a-b)/2$, $B = (a+b)/2$
- (b) $\frac{d}{dt} \left[(t^2 - 1) \frac{dy}{dt} \right] - \alpha(\alpha + 1)y = 0$ onde $\alpha = 1/(m+2)$, desde que $1/(m+2)$
4. $u_1(x) = 1 + \sum_{m=1}^{\infty} (-1)^m 2^m \frac{\alpha(\alpha-2) \cdots (\alpha-2m+2)}{(2m)!} x^{2m}$ para todo x ;

$$u_2(x) = x + \sum_{m=1}^{\infty} (-1)^m 2^m \frac{(\alpha-1)(\alpha-3)\cdots(\alpha-2m+1)}{(2m+1)!} x^{2m+1} \text{ para todo } x$$

$$5. u_1(x) = 1 + \sum_{m=1}^{\infty} \frac{(-1)^m}{(3m+2)(3m-1)\cdots 8 \cdot 5} x^{3m} \text{ para todo } x;$$

$$u_2(x) = x^{-2} \left(1 + \sum_{n=1}^{\infty} \frac{(-1)^n}{3^n n!} x^{3n} \right) \text{ para todo } x \neq 0.$$

$$6. y = x^2 \left(\frac{1}{6} + \sum_{n=1}^{\infty} \frac{(\alpha-2)(\alpha-3)\cdots(\alpha-n-1)}{n!(n+3)!} x^n \right) \text{ para todo } x$$

$$11. (b) f(x) = \frac{1}{6} P_0(x) + \frac{4}{7} P_2(x) + \frac{8}{35} P_4(x)$$

$$15. (b) \frac{2n}{4n^2 - 1}$$

6.24. Exercícios (pg. 212)

$$5. J_{-\frac{3}{2}}(x) = -\left(\frac{2}{\pi x}\right)^{\frac{1}{2}} \left(\frac{\cos x}{x} + \sin x \right)$$

$$9. (a) y = x^{\frac{1}{2}} [c_1 J_{\frac{1}{2}}(\frac{2}{3} x^{\frac{3}{2}}) + c_2 J_{-\frac{1}{2}}(\frac{2}{3} x^{\frac{3}{2}})]$$

$$(b) y = x^{\frac{1}{2}} [c_1 J_{\frac{1}{2}}(\frac{1}{2} x^2) + c_2 J_{-\frac{1}{2}}(\frac{1}{2} x^2)]$$

$$(c) y = x^{\frac{1}{2}} [c_1 J_{\alpha}(2\alpha x^{1+m/2}) + c_2 J_{-\alpha}(2\alpha x^{1+m/2})], \text{ onde } \alpha = 1/(m+2), \text{ desde que } 1/(m+2) \text{ não seja inteiro; caso contrário, substituir a } J \text{ correspondente por } K.$$

$$(d) y = x^{\frac{1}{2}} [c_1 J_{\alpha}(\frac{1}{2} x^2) + c_2 J_{-\alpha}(\frac{1}{2} x^2)], \text{ onde: } \alpha = \sqrt{2}/8$$

$$10. y = g_{\alpha} \text{ satisfaz } x^2 y'' + (1-2c)xy' + (a^2 b^2 x^{2b} + c^2 - \alpha^2 b^2)y = 0$$

$$(a) y = x^{-\frac{5}{2}} [c_1 J_5(2x^{\frac{1}{2}}) + c_2 K_5(2x^{\frac{1}{2}})]$$

$$(b) y = x^{-\frac{5}{2}} [c_1 J_{\frac{5}{2}}(x) + c_2 J_{-\frac{5}{2}}(x)]$$

$$(c) y = x^{-\frac{5}{2}} [c_1 J_1(\frac{2}{3} x^{\frac{3}{2}}) + c_2 K_1(\frac{2}{3} x^{\frac{3}{2}})]$$

$$(d) y = x [c_1 J_0(2x^{\frac{1}{2}}) + c_2 K_0(2x^{\frac{1}{2}})]$$

$$11. a = 2, c = 0$$

$$12. y = \sum_{n=0}^{\infty} \frac{(-1)^n}{(n!)^2} x^n; y = J_0(2x^{\frac{1}{2}}) \text{ if } x > 0$$

$$13. b = (p_0' - a_0)/a_0, c = q_0/a_0$$

$$14. y = x^{\frac{1}{2}}$$

$$15. t = 1: y = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{n!}{(2n)!} (2x)^n$$

$$t = \frac{1}{2}: y = x^{\frac{1}{2}} \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \left(\frac{x}{2}\right)^n = x^{\frac{1}{2}} e^{-x/2}$$

$$16. u_0(x) = \cos x; u_1(x) = \frac{1}{2} - \frac{1}{6} \cos x - \frac{1}{6} \cos 2x$$

Capítulo 7

7.4. Exercícios (pg. 220)

$$3. (b) (P^k)' = \sum_{m=0}^{k-1} P^m P' P^{k-1-m}$$

7.12. Exercícios (pg. 230)

$$1. (a) A^{-1} = 2I - A, \quad A^n = nA - (n-1)I$$

$$(b) e^{tA} = e^t(1-t)I + te^tA = e^t \begin{bmatrix} 1 & 0 \\ t & 1 \end{bmatrix}$$

$$2. (a) A^{-1} = \frac{3}{2}I - \frac{1}{2}A, \quad A^n = (2^n - 1)A - (2^n - 2)I$$

$$(b) e^{tA} = (2e^t - e^{2t})I + (e^{2t} - e^t)A = \begin{bmatrix} e^t & 0 \\ e^{2t} - e^t & e^{2t} \end{bmatrix}$$

$$3. (a) A^{-1} = A, \quad A^n = \frac{1 + (-1)^n}{2} I + \frac{1 - (-1)^n}{2} A$$

$$(b) e^{tA} = (\operatorname{ch} t)I + (\operatorname{sh} t)A = \begin{bmatrix} \operatorname{ch} t & \operatorname{sh} t \\ \operatorname{sh} t & \operatorname{ch} t \end{bmatrix}$$

$$4. (a) A^{-1} = A, \quad A^n = \frac{1 + (-1)^n}{2} I + \frac{1 - (-1)^n}{2} A$$

$$(b) e^{tA} = (\operatorname{ch} t)I + (\operatorname{sh} t)A = \begin{bmatrix} e^{-t} & 0 \\ 0 & e^t \end{bmatrix}$$

$$5. (b) e^{tA} = e^{at} \begin{bmatrix} \cos bt & \operatorname{sen} bt \\ -\operatorname{sen} bt & \cos bt \end{bmatrix}$$

$$7. e^{A(t)} = I + (e-1)A(t); \quad (e^{A(t)})' = (e-1)A'(t) = \begin{bmatrix} 0 & e-1 \\ 0 & 0 \end{bmatrix};$$

$$e^{A(t)}A'(t) = \begin{bmatrix} 0 & e \\ 0 & 0 \end{bmatrix}; \quad A'(t)e^{A(t)} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

$$8. (a) A^n = O \quad \text{se } n \geq 3$$

$$(b) e^{tA} = I + tA + \frac{1}{2}t^2A^2 = \begin{bmatrix} 1 & t & t + \frac{1}{2}t^2 \\ 0 & 1 & t \\ 0 & 0 & 1 \end{bmatrix}$$

$$9. (a) A^n = A \quad \text{se } n \geq 1$$

$$(b) e^{tA} = I + (e^t - 1)A = \begin{bmatrix} 1 & e^t - 1 & e^t - 1 \\ 0 & e^t & e^t - 1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$10. (a) A^3 = 4A^2 - 5A + 2I; \quad A^n = \begin{bmatrix} 2^n & 0 & 0 \\ 0 & 1 & 0 \\ 0 & n & 1 \end{bmatrix}$$

$$(b) e^{tA} = \begin{bmatrix} e^{2t} & 0 & 0 \\ 0 & e^t & 0 \\ 0 & te^t & e^t \end{bmatrix}$$

$$11. e^{tA} = I + tA + \frac{1}{2}t^2A^2$$

$$13. e^A e^B = \begin{bmatrix} e^2 & -(e-1)^2 \\ 0 & 1 \end{bmatrix}; \quad e^B e^A = \begin{bmatrix} e^2 & (e-1)^2 \\ 0 & 1 \end{bmatrix}; \quad e^{A+B} = \begin{bmatrix} e^2 & 0 \\ 0 & 1 \end{bmatrix}$$

7.15. Exercícios (pg. 238)

$$1. e^{tA} = \frac{1}{2}(3e^t - e^{3t})I + \frac{1}{2}(e^{3t} - e^t)A$$

$$2. e^{tA} = (\operatorname{ch} \sqrt{5} t)I + \frac{1}{\sqrt{5}}(\operatorname{sh} \sqrt{5} t)A$$

$$3. e^{tA} = \frac{1}{2}e^t\{(t^2 - 2t + 2)I + (-2t^2 + 2t)A + t^2A^2\}$$

$$4. e^{tA} = (3e^{-t} - 3e^{-2t} + e^{-3t})I + (\frac{5}{2}e^{-t} - 4e^{-2t} + \frac{3}{2}e^{-3t})A + (\frac{1}{2}e^{-t} - e^{-2t} + \frac{1}{2}e^{-3t})A^2$$

$$5. e^{tA} = (4e^t - 3e^{2t} + 2te^{2t})I + (4e^{2t} - 3te^{2t} - 4e^t)A + (e^t - e^{2t} + te^{2t})A^2$$

$$6. e^{tA} = (4e^t - 6e^{2t} + 4e^{3t} - e^{4t})I + (-\frac{1}{3}e^t + \frac{1}{3}e^{2t} - 7e^{3t} + \frac{1}{6}e^{4t})A + (\frac{3}{2}e^t - 4e^{2t} + \frac{7}{2}e^{3t} - e^{4t})A^2 + (-\frac{1}{6}e^t + \frac{1}{2}e^{2t} - \frac{1}{2}e^{3t} + \frac{1}{6}e^{4t})A^3$$

$$7. (b) e^{tA} = \frac{1}{6}e^{\lambda t}\{(6 - 6\lambda t + 3\lambda^2 t^2 - \lambda^3 t^3)I + (6t - 6\lambda t^2 + 3\lambda^2 t^3)A + (3t^2 - 3\lambda t^3)A^2 + t^3A^3\}$$

$$8. y_1 = c_1 \operatorname{ch} \sqrt{5} t + \frac{c_1 + 2c_2}{\sqrt{5}} \operatorname{sh} \sqrt{5} t, \quad y_2 = c_2 \operatorname{ch} \sqrt{5} t + \frac{2c_1 - c_2}{\sqrt{5}} \operatorname{sh} \sqrt{5} t$$

$$9. y_1 = e^t(\cos 3t - \operatorname{sen} 3t), \quad y_2 = e^t(\cos 3t - 3 \operatorname{sen} 3t)$$

$$10. y_1 = e^{2t} + 4te^{2t}, \quad y_2 = -2e^t + e^{2t} + 4te^{2t}, \quad y_3 = -2e^t + 4e^{2t}$$

$$11. y_1 = c_1 e^{2t}, \quad y_2 = c_2 e^t, \quad y_3 = (c_2 t + c_3)e^t$$

$$12. y_1 = 3e^{-t} - 3e^{-2t} + e^{-3t}, \quad y_2 = -3e^{-t} + 6e^{-2t} - 3e^{-3t}, \quad y_3 = 3e^{-t} - 12e^{-2t} + 9e^{-3t}$$

$$13. y_1 = e^{5t} + 7e^{-3t}, \quad y_2 = 2e^{5t} - 2e^{-3t}, \quad y_3 = -e^{5t} + e^{-3t}$$

$$14. y_1 = -\frac{1}{2}e^t + e^{2t} + \frac{1}{2}e^{3t}, \quad y_2 = e^{2t} + e^{3t}, \quad y_3 = e^{3t}, \quad y_4 = e^{4t}$$

$$15. y_1 = 2e^{2t} - 1, \quad y_2 = 2e^{2t} - t - 2, \quad y_3 = 2e^{2t}, \quad y_4 = e^{2t}$$

7.17. Exercícios (pg. 241)

$$2. (c) y_1 = (b-1)e^x + 2(c+1-b)xe^x + 1, \quad y_2 = ce^x + 2(c+1-b)xe^x$$

$$4. y_1 = -\frac{1}{3}e^t - \frac{1}{6}e^{4t} + \frac{1}{2}e^{2t}, \quad y_2 = \frac{2}{3}e^t - \frac{1}{6}e^{4t} + \frac{1}{2}e^{2t}$$

$$5. (a) B_0 = B, \quad B_1 = AB, \quad B_2 = \frac{1}{2!}A^2B, \dots, \quad B_m = \frac{1}{m!}A^mB$$

$$(b) B = -m!(A^{-1})^{m+1}C$$

$$6. (a) Y(t) = \left(I + tA + \frac{1}{2}t^2A^2 + \frac{1}{6}t^3A^3 \right) B, \quad \text{onde } B = -6A^{-4}C = -\frac{3}{128} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Isto dá-nos a solução particular $y_1 = y_2 = -\frac{3}{128} - \frac{3}{32}t - \frac{3}{16}t^2 - \frac{1}{4}t^3$

$$(b) \quad y_1 = y_2 = -\frac{3}{128} - \frac{3}{32}t - \frac{3}{16}t^2 - \frac{1}{4}t^3 + \frac{131}{128}e^{4t}$$

$$7. \quad E = B, \quad F = \frac{1}{\alpha}(AB + C)$$

$$8. \quad (a) \quad y_1 = -\cos 2t - \frac{1}{2}\sin 2t, \quad y_2 = -\frac{1}{2}\sin 2t$$

$$(b) \quad y_1 = 2 \cosh 2t + \frac{1}{2} \sinh 2t - \cos 2t - \frac{1}{2} \sin 2t, \quad y_2 = \cosh 2t + \frac{1}{2} \sinh 2t - \frac{1}{2} \sin 2t.$$

$$9. \quad y_1(x) = e^{2x} + e^{3x} - e^x, \quad y_2(x) = -2e^{2x} - e^{3x} + 3e^x$$

$$10. \quad y_1(x) = \frac{4}{25}e^x - \frac{3}{8}e^{2x} + (c_1 - \frac{11}{90}e^{-4x} + (\frac{11}{30} - c_1 - c_2)xe^{-4x},$$

$$y_2(x) = \frac{1}{25}e^x + \frac{7}{8}e^{2x} + (c_2 - \frac{211}{900})e^{-4x} + (c_1 + c_2 - \frac{11}{30})xe^{-4x}$$

$$11. \quad y_1(x) = e^{-4x}(2 \cos x + \sin x) + \frac{31}{8}e^x - \frac{9}{17}, \quad y_2(x) = e^{-4x}(\sin x + 3 \cos x) - \frac{2}{13}e^x + \frac{6}{17}$$

$$12. \quad y_1(x) = e^{-x}(x^2 + 2x + 3) + x^2 - 3x + 3, \quad y_2(x) = e^{-x}(-2x - 2) + x,$$

$$y_3(x) = 2e^{-x} + x - 1$$

7.20. Exercícios (pg. 249)

$$4. \quad (c) \quad Y(x) = e^x e^{\frac{1}{2}x^2} AB$$

$$5. \quad \text{Se } A(x) = \sum_{k=0}^{\infty} x^k A_k, \text{ então } Y(x) = B + xC + \sum_{k=2}^{\infty} x^k B_k,$$

$$\text{onde } (k+2)(k+1)B_{k+2} = \sum_{r=0}^k A_r B_{k-r} \text{ para } k \geq 0.$$

7.24. Exercícios (pg. 259)

$$1. \quad (a) \quad Y(x) = e^x$$

$$(b) \quad Y_n(x) = 2e^x - \sum_{k=0}^n \frac{x^k}{k!} \text{ se } n \text{ é ímpar; } Y_n(x) = \sum_{k=0}^n \frac{x^k}{k!} \text{ se } n \text{ é par.}$$

$$2. \quad Y_3(x) = \frac{x^2}{2} + \frac{x^5}{20} + \frac{x^8}{160} + \frac{x^{11}}{4400}$$

$$3. \quad Y_3(x) = x + \frac{x^4}{4} + \frac{x^7}{14} + \frac{x^{10}}{160}$$

$$4. \quad Y_3(x) = \frac{x^3}{3} + \frac{4x^7}{63} + \frac{8x^9}{405} + \frac{184x^{11}}{51975} + \frac{4x^{13}}{12285}$$

$$5. \quad (a) \quad Y_2(x) = 1 + x + x^2 + \frac{2x^3}{3} + \frac{x^4}{6} + \frac{2x^5}{15} + \frac{x^7}{63}$$

$$(b) \quad M = 2; \quad c = \frac{1}{2}$$

$$(c) \quad Y(x) = 1 + x + x^2 + \frac{4x^3}{3} + \frac{7x^4}{6} + \frac{6x^5}{5} + \dots$$

$$6. \quad (a) \quad Y_4(x) = x + \frac{x^3}{3} + \frac{2x^5}{15} + \frac{17x^7}{315} + \frac{38x^9}{2835} + \frac{134x^{11}}{51975} + \frac{4x^{13}}{12285} + \frac{x^{15}}{59535}$$

$$(d) \quad Y(x) = \tan x = x + \frac{x^3}{3} + \frac{2x^5}{15} + \frac{17x^7}{315} + \frac{62x^9}{2835} + \dots \text{ para } |x| < \frac{\pi}{2}$$

$$8. \quad Y_3(x) = 2 + x^2 + x^3 + \frac{3x^5}{20} + \frac{x^6}{10}; \quad Z_3(x) = 3x^2 + \frac{3x^4}{4} + \frac{6x^5}{5} + \frac{3x^7}{28} + \frac{3x^8}{40}$$

$$9. \quad Y_3(x) = 5 + x + \frac{x^4}{12} + \frac{x^6}{6} + \frac{2x^7}{63} + \frac{x^9}{72};$$

$$Z_3(x) = 1 + \frac{x^3}{3} + x^5 + \frac{2x^6}{9} + \frac{x^8}{8} + \frac{11x^9}{324} + \frac{7x^{11}}{264}$$

$$10. \quad (d) \quad Y_n(x) = 0; \quad \lim_{n \rightarrow \infty} Y_n(x) = 0$$

$$(e) \quad Y_n(x) = \begin{cases} x^2 & \text{se } x \geq 0 \\ -x^2 & \text{se } x \leq 0 \end{cases}; \quad \lim_{n \rightarrow \infty} Y_n(x) = \begin{cases} x^2 & \text{si } x \geq 0 \\ -x^2 & \text{si } x \leq 0 \end{cases}$$

$$(f) \quad Y_n(x) = \frac{2x^n}{3^n}; \quad \lim_{n \rightarrow \infty} Y_n(x) = 0$$

$$(g) \quad Y_n(x) = \begin{cases} x^2 & \text{se } x \geq 0 \\ -x^2 & \text{se } x \leq 0 \end{cases}; \quad \lim_{n \rightarrow \infty} Y_n(x) = \begin{cases} x^2 & \text{si } x \geq 0 \\ -x^2 & \text{si } x \leq 0 \end{cases}$$

Capítulo 8

8.3. Exercícios (pg. 276)

2. Todos abertos excepto (d), (e), (h) e (j)
3. Todos abertos excepto (d)
5. (c) Um exemplo é a família de todas as 2-esferas $B(O; 1/k)$, onde $k = 1, 2, 3, \dots$
6. (a) Aberto e fechado (b) Aberto e fechado (c) Fechado (d) Aberto (e) Fechado (f) Nem aberto nem fechado (g) Fechado (h) Nem aberto nem fechado (i) Fechado (j) Fechado (k) Nem aberto nem fechado (l) Fechado
8. (e) Um exemplo é a união de todos os conjuntos da forma $S_k = \{x \mid \|x\| \leq 1 - 1/k\}$ para $k = 1, 2, 3, \dots$. A sua união é a esfera aberta $B(O; 1)$
10. Não.

8.5. Exercícios (pg. 282)

1. (a) Todo (x, y)
 (b) Todo $(x, y) \neq (0, 0)$
 (c) Todo (x, y) com $y \neq 0$
 (d) Todo (x, y) com $y \neq 0$ e $\frac{x^2}{y} \neq \frac{\pi}{2} + k\pi$ ($k = 0, 1, 2, \dots$)
 (e) Todo (x, y) com $x \neq 0 \neq (0, 0)$
 (f) Todo (x, y)
 (g) Todo (x, y) com $xy \neq 1$
 (h) Todo $(x, y) \neq (0, 0)$

- (i) Todo $(x, y) \neq (0, 0)$.
 (j) Todo (x, y) com $y \neq 0$ e $0 \leq x \leq y$ ou $y \leq x \leq 0$
5. $\lim_{y \rightarrow 0} f(x, y)$ não existe se $x \neq 0$
 6. $(1 - m^2)/(1 + m^2)$; Não
 7. $y = \frac{1}{2}x^2$; f não é contínua em $(0, 0)$
 8. $f(0, 0) = 1$

8.9. Exercícios (pg. 287)

1. $f'(x; y) = a \cdot y$
 2. (a) $f'(x; y) = 4 \|x\|^2 x \cdot y$
 (b) Todos os pontos da recta $2x + 3y = \frac{1}{2}$
 (c) Todos os pontos do plano $x + 2y + 3z = 0$
 3. $f'(x; y) = x \cdot T(y) + y \cdot T(x)$
 4. $\frac{\partial f}{\partial x} = 2x + y^3 \cos(xy)$; $\frac{\partial f}{\partial y} = 2y \sin(xy) + xy^2 \cos(xy)$
 5. $\frac{\partial f}{\partial x} = x/(x^2 + y^2)^{3/4}$; $\frac{\partial f}{\partial y} = y/(x^2 + y^2)^{3/4}$
 6. $\frac{\partial f}{\partial x} = y^2/(x^2 + y^2)^{3/2}$; $\frac{\partial f}{\partial y} = -xy/(x^2 + y^2)^{3/2}$
 7. $\frac{\partial f}{\partial x} = -2y/(x - y)^2$; $\frac{\partial f}{\partial y} = 2x/(x - y)^2$
 8. $D_k f(x) = a_k$, onde $a = (a_1, \dots, a_n)$
 9. $D_k f(x) = 2 \sum_{j=1}^n a_{kj} x_j$
 10. $\frac{\partial f}{\partial x} = 4x^3 - 8xy^2$; $\frac{\partial f}{\partial y} = 4y^3 - 8x^2y$
 11. $\frac{\partial f}{\partial x} = \frac{2x}{x^2 + y^2}$; $\frac{\partial f}{\partial y} = \frac{2y}{x^2 + y^2}$
 12. $\frac{\partial f}{\partial x} = -\frac{2x}{y} \sin(x^2)$; $\frac{\partial f}{\partial y} = -\frac{1}{y^2} \cos(x^2)$
 13. $\frac{\partial f}{\partial x} = \frac{2x}{y} \sec^2 \frac{x^2}{y}$; $\frac{\partial f}{\partial y} = -\frac{x^2}{y^2} \sec^2 \frac{x^2}{y}$
 14. $\frac{\partial f}{\partial x} = -\frac{y}{x^2 + y^2}$; $\frac{\partial f}{\partial y} = \frac{1}{x^2 + y^2}$
 15. $\frac{\partial f}{\partial x} = \frac{1 + y^2}{1 + x^2 + y^2 + x^2 y^2}$; $\frac{\partial f}{\partial y} = \frac{1 + x^2}{1 + x^2 + y^2 + x^2 y^2}$
 16. $\frac{\partial f}{\partial x} = y^2 x^{y^2-1}$; $\frac{\partial f}{\partial y} = 2yx^{y^2} \log x$

17. $\frac{\partial f}{\partial x} = -\frac{1}{2\sqrt{x(y-x)}}; \frac{\partial f}{\partial y} = \frac{\sqrt{x}}{2y\sqrt{y-x}}$
18. $n = -\frac{3}{2}$
19. $a = b = 1$
22. (b) Um exemplo é $f(x) = x \cdot y$, com y um vector dado não nulo

8.14. Exercícios (pg. 295)

1. (a) $(2x + y^3 \cos xy)i + (2y \sin xy + xy^2 \cos xy)j$
 (b) $e^x \cos yj - e^x \sin yj$
 (c) $2xy^3z^4i + 3x^2y^2z^4j + 4x^2y^3z^3k$
 (d) $2xi - 2yj + 4zk$
 (e) $\frac{2x}{x^2 + 2y^2 - 3z^2}i + \frac{4y}{x^2 + 2y^2 - 3z^2}j - \frac{6z}{x^2 + 2y^2 - 3z^2}k$
 (f) $y^zx^{y^z-1}i + zy^{z-1}x^{y^z} \log xj + y^zx^{y^z} \log x \log yk$
2. (a) $-2/\sqrt{6}$
 (b) $1/\sqrt{6}$
3. $(1, 0)$, na direcção de i ; $(-1, 0)$, na direcção de $-i$
4. $2i + 2j; \frac{1}{5}i$
5. $(a, b, c) = (6, 24, -8)$ ou $(-6, -24, 8)$
6. O conjunto de pontos (x, y) da recta $5x - 3y = 6$; $\Delta f(a) = 5i - 3j$
8. (c) Sim
 (d) $f(x, y, z) = \frac{1}{2}(x^2 + y^2 + z^2)$
11. (b) implica (a) e (c); (d) implica (a), (b) e (c); (f) implica (a)

8.17. Exercícios (pg. 301)

1. (b) $F''(t) = \frac{\partial^2 f}{\partial x^2} [X'(t)]^2 + 2 \frac{\partial^2 f}{\partial x \partial y} X'(t) Y'(t) + \frac{\partial^2 f}{\partial y^2} [Y'(t)]^2 + \frac{\partial f}{\partial x} X''(t) + \frac{\partial f}{\partial y} Y''(t)$
2. (a) $F'(t) = 4t^3 + 2t$; $F''(t) = 12t^2 + 2$
 (b) $F'(t) = (2 \cos^2 t - 1)e^{\cos t \sin t} \cos(\cos t \sin^2 t) + (3 \sin^3 t - 2 \sin t)e^{\cos t \sin t} \sin(\cos t \sin^2 t)$;
 $F''(t) = (5 \cos^6 t - 3 \cos^4 t - 4 \cos^3 t - \cos^2 t - 4 \cos t)e^{\cos t \sin t} \cos(\cos t \sin^2 t)$
 $+ (14 \sin^3 t - 12 \sin^5 t - 4 \sin t + 7 \cos t - 9 \cos^3 t)e^{\cos t \sin t} \sin(\cos t \sin^2 t)$
- (c) $F'(t) = \frac{2e^{2t} \exp(e^{2t})}{1 + \exp(e^{2t})} + \frac{2e^{-2t} \exp(e^{-2t})}{1 + \exp(e^{-2t})}$, onde $\exp(u) = e^u$;
 $F''(t) = \frac{4[1 + e^{2t} + \exp(e^{2t})]e^{2t} \exp(e^{2t})}{[1 + \exp(e^{2t})]^2} - \frac{4[1 + e^{-2t} + \exp(e^{-2t})]e^{-2t} \exp(e^{-2t})}{[1 + \exp(e^{-2t})]^2}$
3. (a) $-\frac{2}{3}$
 (b) $x^2 - y^2$
 (c) 0

4. (a) $(1 + 3x^2 + 3y^2)(xi + yj) - (x^2 + y^2)^{1/2}k$, ou qualquer produto deste por um escalar
 (b) $\cos \theta = -[1 + (1 + 3(x^2 + y^2))^2]^{-1/2}$; $\cos \theta \rightarrow -\frac{1}{2}\sqrt{2}$ quando $(x, y, z) \rightarrow (0, 0, 0)$
 5. $U(x, y) = \frac{1}{2} \log(x^2 + y^2)$; $V(x, y) = \arctg(y/x)$
 6. (b) Não
 8. $x/x_0 + y/y_0 + z/z_0 = 3$
 9. $x + y + 2z = 4$, $x - y - z = -1$
 10. $c = \pm\sqrt{3}$

8.22. Exercícios (pg. 309)

1. (b) $\frac{\partial f}{\partial x} = -2x \sin(x^2 + y^2) \cos[\cos(x^2 + y^2)]e^{\sin[\cos(x^2 + y^2)]}$
2. $\frac{\partial F}{\partial x} = \frac{1}{2} \frac{\partial f}{\partial u} + \frac{1}{2} \frac{\partial f}{\partial v}$; $\frac{\partial F}{\partial y} = -\frac{1}{2} \frac{\partial f}{\partial u} + \frac{1}{2} \frac{\partial f}{\partial v}$
3. (a) $\frac{\partial F}{\partial s} = \frac{\partial f}{\partial x} \frac{\partial X}{\partial s} + \frac{\partial f}{\partial y} \frac{\partial Y}{\partial s}$; $\frac{\partial F}{\partial t} = \frac{\partial f}{\partial x} \frac{\partial X}{\partial t} + \frac{\partial f}{\partial y} \frac{\partial Y}{\partial t}$
 (c) $\frac{\partial^2 F}{\partial s \partial t} = \frac{\partial^2 f}{\partial x^2} \frac{\partial X}{\partial s} \frac{\partial X}{\partial t} + \frac{\partial^2 f}{\partial x \partial y} \left(\frac{\partial X}{\partial s} \frac{\partial Y}{\partial t} + \frac{\partial X}{\partial t} \frac{\partial Y}{\partial s} \right) + \frac{\partial^2 f}{\partial y^2} \frac{\partial Y}{\partial s} \frac{\partial Y}{\partial t} + \frac{\partial f}{\partial x} \frac{\partial^2 X}{\partial s \partial t} + \frac{\partial f}{\partial y} \frac{\partial^2 Y}{\partial s \partial t}$
4. (a) $\frac{\partial F}{\partial s} = \frac{\partial f}{\partial x} + t \frac{\partial f}{\partial y}$; $\frac{\partial F}{\partial t} = \frac{\partial f}{\partial x} + s \frac{\partial f}{\partial y}$; $\frac{\partial^2 F}{\partial s^2} = \frac{\partial^2 f}{\partial x^2} + 2t \frac{\partial^2 f}{\partial x \partial y} + t^2 \frac{\partial^2 f}{\partial y^2}$;
 $\frac{\partial^2 F}{\partial t^2} = \frac{\partial^2 f}{\partial x^2} + 2s \frac{\partial^2 f}{\partial x \partial y} + s^2 \frac{\partial^2 f}{\partial y^2}$; $\frac{\partial^2 F}{\partial s \partial t} = \frac{\partial^2 f}{\partial x^2} + (s + t) \frac{\partial^2 f}{\partial x \partial y} + st \frac{\partial^2 f}{\partial y^2} + \frac{\partial f}{\partial y}$
 (b) $\frac{\partial F}{\partial s} = t \frac{\partial f}{\partial x} + \frac{1}{t} \frac{\partial f}{\partial y}$; $\frac{\partial F}{\partial t} = s \frac{\partial f}{\partial x} - \frac{s}{t^2} \frac{\partial f}{\partial y}$; $\frac{\partial^2 F}{\partial s^2} = t^2 \frac{\partial^2 f}{\partial x^2} + 2 \frac{\partial^2 f}{\partial x \partial y} + \frac{1}{t^2} \frac{\partial^2 f}{\partial y^2}$;
 $\frac{\partial^2 F}{\partial t^2} = s^2 \frac{\partial^2 f}{\partial x^2} - 2 \frac{s^2}{t^3} \frac{\partial^2 f}{\partial x \partial y} + \frac{s^2}{t^4} \frac{\partial^2 f}{\partial y^2} + \frac{2s}{t^3} \frac{\partial f}{\partial y}$; $\frac{\partial^2 F}{\partial s \partial t} = st \frac{\partial^2 f}{\partial x^2} - \frac{s}{t^3} \frac{\partial^2 f}{\partial y^2} + \frac{\partial f}{\partial x} - \frac{1}{t^2} \frac{\partial f}{\partial y}$
 (c) $\frac{\partial F}{\partial s} = \frac{1}{2} \frac{\partial f}{\partial x} + \frac{1}{2} \frac{\partial f}{\partial y}$; $\frac{\partial F}{\partial t} = -\frac{1}{2} \frac{\partial f}{\partial x} + \frac{1}{2} \frac{\partial f}{\partial y}$; $\frac{\partial^2 F}{\partial s \partial t} = -\frac{1}{4} \frac{\partial^2 f}{\partial x^2} + \frac{1}{4} \frac{\partial^2 f}{\partial y^2}$;
 $\frac{\partial^2 F}{\partial s^2} = \frac{1}{4} \frac{\partial^2 f}{\partial x^2} + \frac{1}{2} \frac{\partial^2 f}{\partial x \partial y} + \frac{1}{4} \frac{\partial^2 f}{\partial y^2}$; $\frac{\partial^2 F}{\partial t^2} = \frac{1}{4} \frac{\partial^2 f}{\partial x^2} - \frac{1}{2} \frac{\partial^2 f}{\partial x \partial y} + \frac{1}{4} \frac{\partial^2 f}{\partial y^2}$
5. $\frac{\partial^2 \varphi}{\partial r^2} = \cos^2 \theta \frac{\partial^2 f}{\partial x^2} + \cos \theta \sin \theta \left(\frac{\partial^2 f}{\partial x \partial y} + \frac{\partial^2 f}{\partial y \partial x} \right) + \sin^2 \theta \frac{\partial^2 f}{\partial y^2}$;
 $\frac{\partial^2 \varphi}{\partial r \partial \theta} = -r \cos \theta \sin \theta \frac{\partial^2 f}{\partial x^2} + r \cos^2 \theta \frac{\partial^2 f}{\partial x \partial y} - r \sin^2 \theta \frac{\partial^2 f}{\partial y \partial x} + r \cos \theta \sin \theta \frac{\partial^2 f}{\partial y^2}$
 $- \sin \theta \frac{\partial f}{\partial x} + \cos \theta \frac{\partial f}{\partial y}$;

6. $\frac{\partial F}{\partial r} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial r} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial r} + \frac{\partial f}{\partial z} \frac{\partial z}{\partial r}$; $\frac{\partial F}{\partial s} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial s} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial s} + \frac{\partial f}{\partial z} \frac{\partial z}{\partial s}$;
 $\frac{\partial F}{\partial t} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial t} + \frac{\partial f}{\partial z} \frac{\partial z}{\partial t}$
7. (a) $\frac{\partial F}{\partial r} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} + 2 \frac{\partial f}{\partial z}$; $\frac{\partial F}{\partial s} = \frac{\partial f}{\partial x} - 2 \frac{\partial f}{\partial y} + \frac{\partial f}{\partial z}$; $\frac{\partial F}{\partial t} = \frac{\partial f}{\partial x} + 3 \frac{\partial f}{\partial y} - \frac{\partial f}{\partial z}$
 (b) $\frac{\partial F}{\partial r} = 2r \left(\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} + \frac{\partial f}{\partial z} \right)$; $\frac{\partial F}{\partial s} = 2s \left(\frac{\partial f}{\partial x} - \frac{\partial f}{\partial y} - \frac{\partial f}{\partial z} \right)$; $\frac{\partial F}{\partial t} = 2t \left(\frac{\partial f}{\partial x} - \frac{\partial f}{\partial y} + \frac{\partial f}{\partial z} \right)$
8. $\frac{\partial F}{\partial s} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial s} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial s} + \frac{\partial f}{\partial z} \frac{\partial z}{\partial s}$; $\frac{\partial F}{\partial t} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial t} + \frac{\partial f}{\partial z} \frac{\partial z}{\partial t}$
9. (a) $\frac{\partial F}{\partial s} = 2s \frac{\partial f}{\partial x} + 2s \frac{\partial f}{\partial y} + 2t \frac{\partial f}{\partial z}$; $\frac{\partial F}{\partial t} = 2t \frac{\partial f}{\partial x} - 2t \frac{\partial f}{\partial y} + 2s \frac{\partial f}{\partial z}$
 (b) $\frac{\partial F}{\partial s} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} + \frac{\partial f}{\partial z}$; $\frac{\partial F}{\partial t} = \frac{\partial f}{\partial x} - \frac{\partial f}{\partial y} + \frac{\partial f}{\partial z}$
10. $\frac{\partial F}{\partial r} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial r} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial r}$; $\frac{\partial F}{\partial s} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial s} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial s}$; $\frac{\partial F}{\partial t} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial t}$
11. (a) $\frac{\partial F}{\partial r} = \frac{\partial f}{\partial x}$; $\frac{\partial F}{\partial s} = \frac{\partial f}{\partial x}$; $\frac{\partial F}{\partial t} = \frac{\partial f}{\partial y}$
 (b) $\frac{\partial F}{\partial r} = \frac{\partial f}{\partial x} + 2r \frac{\partial f}{\partial y}$; $\frac{\partial F}{\partial s} = \frac{\partial f}{\partial x} + 2s \frac{\partial f}{\partial y}$; $\frac{\partial F}{\partial t} = \frac{\partial f}{\partial x} + 2t \frac{\partial f}{\partial y}$
 (c) $\frac{\partial F}{\partial r} = \frac{1}{s} \frac{\partial f}{\partial x}$; $\frac{\partial F}{\partial s} = \frac{-r}{s^2} \frac{\partial f}{\partial x} + \frac{1}{t} \frac{\partial f}{\partial y}$; $\frac{\partial F}{\partial t} = \frac{-s}{t^2} \frac{\partial f}{\partial y}$
13. (a) $f(x, y, z) = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$, mais um vector constante qualquer
 (b) $f(x, y, z) = P(x)\mathbf{i} + Q(y)\mathbf{j} + R(z)\mathbf{k}$, onde P, Q, R são três funções quaisquer que verificam $P' = p, Q' = q, R' = r$
14. (a) $Df(x, y) = \begin{bmatrix} e^{x+2y} & 2e^{x+2y} \\ 2 \cos(y+2x) & \cos(y+2x) \end{bmatrix}$; $Dg(u, v, w) = \begin{bmatrix} 1 & 4v & 9w^2 \\ -2u & 2 & 0 \end{bmatrix}$
 (b) $h(u, v, w) = e^{u+2v^2+3w^2+4v-2u^2} \mathbf{i} + \sin(2v - u^2 + 2u + 4v^2 + 6w^3) \mathbf{j}$
 (c) $Dh(1, -1, 1) = \begin{bmatrix} -3 & 0 & 9 \\ 0 & -6 \cos 9 & 18 \cos 9 \end{bmatrix}$
15. (a) $Df(x, y, z) = \begin{bmatrix} 2x & 1 & 1 \\ 2 & 1 & 2z \end{bmatrix}$; $Dg(u, v, w) = \begin{bmatrix} v^2 w^2 & 2uvw^2 & 2uv^2 w \\ 0 & w^2 \cos v & 2w \sin v \\ 2ue^v & u^2 e^v & 0 \end{bmatrix}$
 (b) $h(u, v, w) = (u^2 v^4 w^4 + w^2 \sin v + u^2 e^v) \mathbf{i} + (2uv^2 w^2 + w^2 \sin v + u^4 e^{2v}) \mathbf{j}$

$$(c) \quad Dh(u, 0, w) = \begin{bmatrix} 2u & w^2 + u^2 & 0 \\ 4u^3 & w^2 + 2u^4 & 0 \end{bmatrix}$$

8.24. Exercícios variados (pg. 315)

- Um exemplo: $f(x, y) = 3x$ quando $x = y$, $f(x, y) = 0$ para os restantes valores de x e y
- $D_1 f(0, 0) = 0$; $D_2 f(0, 0) = -1$; $D_{2,1} f(0, 0) = 0$; $D_{1,2} f(0, 0)$ não existe
- (a) Se $a = (a_1, a_2)$, então $f'(0, a) = a_2^2/a_1^2$ se $a_1 \neq 0$, se $f'(0; a) = 0$ se $a_1 = 0$
(b) A função f não é contínua na origem.

$$4. \quad \frac{\partial f}{\partial x} = \frac{1}{2} e^{-xy} x^{-1/2} y^{1/2}; \quad \frac{\partial f}{\partial y} = \frac{1}{2} e^{-xy} x^{1/2} y^{-1/2}$$

$$\begin{aligned} 5. \quad F'''(t) &= \frac{\partial^3 f}{\partial x^3} [X'(t)]^3 + 3 \frac{\partial^2 f}{\partial x^2 \partial y} [X'(t)]^2 Y'(t) + 3 \frac{\partial^2 f}{\partial x \partial y^2} X'(t) [Y'(t)]^2 \\ &+ \frac{\partial^3 f}{\partial y^3} [Y'(t)]^3 + 3 \frac{\partial^2 f}{\partial x^2} X'(t) X''(t) + 3 \frac{\partial^2 f}{\partial x \partial y} [X''(t) Y'(t) + X'(t) Y''(t)] \\ &+ 3 \frac{\partial^2 f}{\partial y^2} Y'(t) Y''(t) + \frac{\partial f}{\partial x} X'''(t) + \frac{\partial f}{\partial y} Y'''(t), \end{aligned}$$

supondo que as derivadas parciais mistas não dependem da ordem de derivação

- 8
- (a) $\frac{\partial g}{\partial u} = \frac{\partial f}{\partial x} v + \frac{\partial f}{\partial y} u$; $\frac{\partial g}{\partial v} = \frac{\partial f}{\partial x} u - \frac{\partial f}{\partial y} v$; $\frac{\partial^2 g}{\partial u \partial v} = uv \frac{\partial^2 f}{\partial x^2} + (u^2 - v^2) \frac{\partial^2 f}{\partial x \partial y} - uv \frac{\partial^2 f}{\partial y^2} + \frac{\partial f}{\partial x}$ (b) $a = \frac{1}{2}$, $b = -\frac{1}{2}$
- (a) $\varphi'(t) = A'(t) \int_c^{B(t)} f[A(t), y] dy + B'(t) \int_a^{A(t)} f[x, B(t)] dx$
(b) $\varphi'(t) = 2te^{t^2}(2e^{t^2} - e^a - e^c)$
- Uma esfera com centro na origem e raio $\sqrt{2}$
- $f(x) = x^2$

Capítulo 9.

9.3. Exercícios (pg. 322)

- $f(x, y) = \sin(x - \frac{4}{3}y)$
- $f(x, y) = e^{x+5y/2} - 1$
- (a) $u(x, y) = x^2 y^2 e^{xy}$
(b) $v(x, y) = 2 + \log \left| \frac{x}{y} \right|$
- $A = B = C = 1$, $D = -3$; $f(x, y) = \varphi_1(3x + y) + \varphi_2(x - y)$
- $G(x, y) = x - y$

9.8. Exercícios (pg. 340)

- $\partial X/\partial v = (1 + xu)/(x - y)$; $\partial Y/\partial u = (1 - yv)/(x - y)$; $\partial Y/\partial v = (1 + yu)/(y - x)$
- $\partial X/\partial y = -(1 + xu)/(1 + u)$; $\partial V/\partial u = (1 - yv)/(1 + yu)$; $\partial V/\partial y = (1 - x)/(1 + u)$
- $\frac{\partial X}{\partial v} = \frac{\partial(F, G)}{\partial(y, v)} \bigg/ \frac{\partial(F, G)}{\partial(x, y)}$; $\frac{\partial Y}{\partial u} = \frac{\partial(F, G)}{\partial(u, x)} \bigg/ \frac{\partial(F, G)}{\partial(x, y)}$; $\frac{\partial Y}{\partial v} = \frac{\partial(F, G)}{\partial(v, x)} \bigg/ \frac{\partial(F, G)}{\partial(x, y)}$
- $T = \pm \frac{1}{\sqrt{751}} (24i - 4\sqrt{7}j + 3\sqrt{7}k)$
- $2i + j + \sqrt{3}k$, ou qualquer produto não nulo deste por um escalar
- $\partial x/\partial u = 0$, $\partial x/\partial v = \pi/12$
- $n = 2$
- $\partial f/\partial x = -1/(2y + 2z + 1)$; $\partial f/\partial y = -2(y + z)/(2y + 2z + 1)$;
 $\partial^2 f/(\partial x \partial y) = 2/(2y + 2z + 1)^3$
- $\partial^2 z/(\partial x \partial y) = [\sin(x + y) \cos^2(y + z) + \sin(y + z) \cos^2(x + y)]/\cos^3(y + z)$
- $\frac{\partial f}{\partial x} = -\frac{D_1 F + 2x D_2 F}{D_1 F + 2z D_2 F}$; $\frac{\partial f}{\partial y} = -\frac{D_1 F + 2y D_2 F}{D_1 F + 2z D_2 F}$
- $D_1 F = f'[x + g(y)]$; $D_2 F = f'[x + g(y)]g'(y)$; $D_{1,1} F = f''[x + g(y)]$;
 $D_{1,2} F = f''[x + g(y)]g'(y)$; $D_{2,2} F = f''[x + g(y)][g'(y)]^2 + f'[x + g(y)]g''(y)$

9.13. Exercícios (pg. 352)

- Mínimo absoluto em $(0, 1)$
- Ponto sela em $(0, 1)$
- Ponto sela em $(0, 0)$
- Mínimo absoluto em cada ponto da recta $y = x + 1$
- Ponto sela em $(1, 1)$
- Mínimo absoluto em $(1, 0)$
- Ponto sela em $(0, 0)$
- Ponto sela em $(0, 6)$ e em $(x, 0)$ para todo x ; mínimo relativo em $(0, y)$, $0 < y < 6$; máximo relativo em $(2, 3)$ e em $(0, y)$ para $y < 0$ e $y > 6$
- Ponto sela em $(0, 0)$; mínimo relativo em $(1, 1)$
- Ponto sela em $(n\pi + \pi/2, 0)$, com n inteiro
- Mínimo absoluto em $(0, 0)$; ponto sela em $(-\frac{1}{4}, -\frac{1}{4})$
- Mínimo absoluto em $(-\frac{1}{26}, -\frac{1}{26})$; máximo absoluto em $(1, 3)$
- Máximo absoluto em $(\pi/3, \pi/3)$; mínimo absoluto em $(2\pi/3, 2\pi/3)$; máximo relativo em (π, π) ; mínimo relativo em $(0, 0)$; ponto sela em $(0, \pi)$ e $(\pi, 0)$
- Ponto sela em $(1, 1)$
- Máximo absoluto em cada ponto da circunferência $x^2 + y^2 = 1$; mínimo absoluto em $(0, 0)$
- (c) Máximo relativo em $(2, 2)$; não existem mínimos relativos; pontos sela em $(0, 3)$, $(3, 0)$ e $(3, 3)$
- Máximo relativo $\frac{1}{8}$ em $(\frac{1}{2}, \frac{1}{2})$ e $(-\frac{1}{2}, -\frac{1}{2})$; mínimo relativo $-\frac{1}{8}$ em $(\frac{1}{2}, -\frac{1}{2})$ e $(-\frac{1}{2}, \frac{1}{2})$; pontos sela em $(0, 0)$, $(\pm 1, 0)$, e $(0, \pm 1)$; máximo absoluto 1 em $(1, -1)$ e $(-1, 1)$; mínimo absoluto -1 em $(1, 1)$ e $(-1, -1)$
- (a) $a = 1$, $b = -\frac{1}{8}$
(b) $a = 6 \log 2 - 3\pi/2$, $b = \pi - 3 \log 2$

21. Seja $x^* = \frac{1}{n} \sum_{i=1}^n x_i$, $y^* = \frac{1}{n} \sum_{i=1}^n y_i$, $u_i = x_i - x^*$. Então $a = \left(\sum_{i=1}^n y_i u_i \right) / \left(\sum_{i=1}^n u_i^2 \right)$,
e $b = y^* - ax^*$

22. Seja $x^* = \frac{1}{n} \sum_{i=1}^n x_i$, $y^* = \frac{1}{n} \sum_{i=1}^n y_i$, $z^* = \frac{1}{n} \sum_{i=1}^n z_i$, $u_i = x_i - x^*$, $v_i = y_i - y^*$, e faça-
mos $\Delta = \begin{vmatrix} \sum u_i^2 & \sum u_i v_i \\ \sum u_i v_i & \sum v_i^2 \end{vmatrix}$, considerando-se as somas para $i = 1, 2, \dots, n$. Então

$$a = \frac{1}{\Delta} \begin{vmatrix} \sum u_i z_i & \sum u_i v_i \\ \sum v_i z_i & \sum v_i^2 \end{vmatrix}, \quad b = \frac{1}{\Delta} \begin{vmatrix} \sum v_i z_i & \sum u_i v_i \\ \sum u_i z_i & \sum u_i^2 \end{vmatrix}, \quad c = z^* - ax^* - by^*$$

25. Valores próprios 4, 16, 16; mínimo relativo em (1, 1, 1)

9.15. Exercícios (pg. 357)

1. O valor máximo é $\frac{1}{4}$; não tem mínimo

2. O máximo é 2; o mínimo é 1

3. (a) O máximo é $\frac{\sqrt{a^2 + b^2}}{ab}$ em $(b(a^2 + b^2)^{-1/4}, a(a^2 + b^2)^{-1/4})$; o mínimo é $-\frac{\sqrt{a^2 + b^2}}{ab}$ em $(-b(a^2 + b^2)^{-1/4}, -a(a^2 + b^2)^{-1/4})$

(b) O mínimo é $a^2 b^2 / (a^2 + b^2)$ em $\left(\frac{ab^2}{a^2 + b^2}, \frac{a^2 b}{a^2 + b^2} \right)$; não tem máximo

4. O máximo é $1 + \sqrt{2}/2$ nos pontos $(n\pi + \pi/8, n\pi - \pi/8)$, sendo n um inteiro qualquer; o mínimo é $1 - \sqrt{2}/2$ em $(n\pi + 5\pi/8, n\pi + 3\pi/8)$, sendo n um inteiro qualquer

5. O máximo é 3 em $(\frac{1}{3}, -\frac{2}{3}, -\frac{2}{3})$; o mínimo é -3 em $(-\frac{1}{3}, \frac{2}{3}, \frac{2}{3})$

6. (0, 0, 1) e (0, 0, -1)

7. 1

8. (1, 0, 0), (0, 1, 0), (-1, 0, 0), (0, -1, 0)

9. $\frac{a^a b^b c^c}{(a+b+c)^{a+b+c}}$ at $\left(\frac{a}{a+b+c}, \frac{b}{a+b+c}, \frac{c}{a+b+c} \right)$

10. $abc\sqrt{3}/2$

11. $5 \log r + 3 \log \sqrt{3}$

12. $m^2 = \frac{A + C - \sqrt{(A - C)^2 + 4B^2}}{2(AC - B^2)}$

13. $(4 \pm \sqrt{5})/\sqrt{2}$

14. O ângulo é $\pi/3$; largura do funido $c/3$; área máxima $c^2/4\sqrt{3}$

Capítulo 10

10.5. Exercícios (pg. 368)

1. $-\frac{1}{15}$
2. $-2\pi a^2$
3. $\frac{1}{3^5}$
4. $\frac{4}{3}$
5. 0
6. 40
7. $\frac{2a}{6}$
8. $\frac{5}{2}$
9. $-\frac{368}{10}$
10. -2π
11. 0
12. (a) $-2\sqrt{2}\pi$
(b) $-\pi$

10.9. Exercícios (pg. 372)

1. $\frac{23}{6}$
2. $2a^3$
3. $a = (3c/2)^{1/2}$
4. 0
5. $8\pi(\sin\theta - \cos\theta)$
6. $\pi a^3/4$
7. $-\sqrt{2}$
8. $256a^3/15$
9. $2\pi^2 a^3(1 + 2\pi^2)$
10. $[(2 + t_0^2)^{3/2} - 2\sqrt{2}]/3$
12. momento de inércia = $4a^4$
13. $2\pi/3$
14. $\frac{600 - 36\sqrt{2} - 49 \log(9 - 4\sqrt{2})}{64[6\sqrt{2} + \log(3 + 2\sqrt{2})]}$
15. $\bar{x} = \frac{6ab^2}{3a^2 + 4\pi^2 b^2}; \bar{y} = -\frac{6\pi ab^2}{3a^2 + 4\pi^2 b^2}$
16. $I_x = (a^2 + b^2)^{1/2}[\pi a^4 + (4\pi^3 - \pi/2)a^2 b^2 + 32\pi^5 b^4/5]$
 $I_y = (a^2 + b^2)^{1/2}[\pi a^4 + (4\pi^3 + \pi/2)a^2 b^2 + 32\pi^5 b^4/5]$

10.13. Exercícios (pg. 378)

1. Todos menos (f) estão conectados
6. (a) Não conservativo
(b) $(2e^{2\pi} - 5e^\pi - 5\pi - 3)/10$
7. (b) 3
8. $\frac{8}{5}$
10. $4b^2 - 8\pi b + 4$; ocorre um mínimo quando $b = \pi$

10.18. Exercícios (pg. 387)

1. $\varphi(x, y) = \frac{1}{2}(x^2 + y^2) + C$
2. $\varphi(x, y) = x^2 y + C$
3. $\varphi(x, y) = x^2 e^y + xy - y^2 + C$
4. $\varphi(x, y) = x \sin y + y \cos x + (x^2 + y^2)/2 + C$
5. $\varphi(x, y) = x \sin(xy) + C$
6. $\varphi(x, y, z) = (x^2 + y^2 + z^2)/2 + C$
7. $\varphi(x, y, z) = x^2/2 - y^2/2 + xz - yz + C$

8. f não é um gradiente
9. f não é um gradiente
10. f não é um gradiente
11. $\varphi(x, y, z) = y^2 \sin x + xz^3 - 4y + 2z + C$
12. $\varphi(x, y, z) = x + 2x^2y - x^3z^2 + 2y - z^3 + C$
13. (b) $\varphi(x, y) = \frac{ar^{n+1}}{n+1} + C \operatorname{sen} n \neq -1$; $\varphi(x, y) = a \log r + C \operatorname{sen} n = -1$
15. $\varphi(x) = \frac{r^{p+2}}{p+2} + C \operatorname{sen} p \neq -2$; $\varphi(x) = \log r + C \operatorname{sen} p = -2$
16. $\varphi(x) = g(r) + C$

10.20. Exercícios (pg. 392)

1. $x^2/2 + 2xy + y^2/2 = C$
2. $x^2y = C$
3. $x^3/3 - xy - y/2 + (\operatorname{sen} 2y)/4 = C$
4. $\cos 2x \operatorname{sen} 3y = C$
5. $x^3y + 4x^2y^2 - 12e^y + 12ye^y = C$
6. $\int Q(x)e^{\int P(x)dx} dx - ye^{\int P(x)dx} = C$
8. (a) $x + y = Cy^2$
(b) $y^3/x^3 - 3 \log |x| = C$
9. (a) $6(xy)^{3/2} - (y/x)^{3/2} = C$; $(x^5y)^{-1/2}$ é um factor integrante
(b) $x + e^{-x} \operatorname{sen} y = C$; $e^{-x} \cos y$ é um factor integrante
10. $x^3y^4 + x^4y^5 = C$, $10x^3y^4 + x^5y^5 = C$, respectivamente; x^3y^3 é um factor integrante comum

Capítulo 11

11.9. Exercícios (pg. 407)

- | | |
|---------------------------------|--|
| 1. $\frac{1}{3}$ | 7. 6 |
| 2. 1 | 8. $t^{-3}(e^{t^3} - e^t) + t^{-2} - t^{-1}$ |
| 3. $2\sqrt{3} - \frac{3\pi}{8}$ | 10. $\frac{1}{6}$ |
| 4. $\pi^2/4$ | 11. $\frac{1}{6}(\frac{21}{8} - \sqrt{2})$ |
| 5. 2 | 12. $\pi/2$ |
| 6. 2π | 13. $(\log 2)/6$ |

11.15. Exercícios (pg. 417)

- | | |
|--|---|
| 1. $-3\pi/2$ | 5. $\pi^2 - \frac{4\pi}{9}$ |
| 2. $\frac{3}{2} + \cos 1 + \operatorname{sen} 1 - \cos 2 - 2 \operatorname{sen} 2$ | 6. 6 |
| 3. $e - e^{-1}$ | 7. $\frac{3\pi}{8}$ |
| 4. $\frac{7}{8} \log 2$ | 8. (a) $\frac{8}{3}$ (b) 2 (c) 320π |
| 9. $\int_0^1 \left[\int_x^1 f(x, y) dy \right] dx$ | |

Capítulo 10

10.5. Exercícios (pg. 368)

- $-\frac{1}{16}$
- $-2\pi a^2$
- $\frac{1}{36}$
- $\frac{4}{3}$
- 0
- 40
- $\frac{23}{6}$
- $\frac{5}{2}$
- $-\frac{368}{10}$
- -2π
- 0
- (a) $-2\sqrt{2}\pi$
(b) $-\pi$

10.9. Exercícios (pg. 372)

- $\frac{23}{6}$
- $2a^3$
- $a = (3c/2)^{1/2}$
- 0
- $8\pi(\sin \theta - \cos \theta)$
- $\pi a^2/4$
- $-\sqrt{2}$
- $256a^3/15$
- $2\pi^2 a^3(1 + 2\pi^2)$
- $[(2 + t_0^2)^{3/2} - 2\sqrt{2}]/3$
- momento de inércia $= 4a^4$
- $2\pi/3$
- $\frac{600 - 36\sqrt{2} - 49 \log(9 - 4\sqrt{2})}{64[6\sqrt{2} + \log(3 + 2\sqrt{2})]}$
- $\bar{x} = \frac{6ab^2}{3a^2 + 4\pi^2 b^2}; \bar{y} = -\frac{6\pi ab^2}{3a^2 + 4\pi^2 b^2}$
- $I_x = (a^2 + b^2)^{1/2}[\pi a^4 + (4\pi^3 - \pi/2)a^2 b^2 + 32\pi^5 b^4/5]$
 $I_y = (a^2 + b^2)^{1/2}[\pi a^4 + (4\pi^3 + \pi/2)a^2 b^2 + 32\pi^5 b^4/5]$

10.13. Exercícios (pg. 378)

- Todos menos (f) estão conectados
- (a) Não conservativo
(b) $(2e^{2\pi} - 5e^\pi - 5\pi - 3)/10$
- (b) 3
- $\frac{8}{5}$
- $4b^2 - 8\pi b + 4$; ocorre um mínimo quando $b = \pi$

10.18. Exercícios (pg. 387)

- $\varphi(x, y) = \frac{1}{2}(x^2 + y^2) + C$
- $\varphi(x, y) = x^2 y + C$
- $\varphi(x, y) = x^2 e^y + xy - y^2 + C$
- $\varphi(x, y) = x \sin y + y \cos x + (x^2 + y^2)/2 + C$
- $\varphi(x, y) = x \sin(xy) + C$
- $\varphi(x, y, z) = (x^2 + y^2 + z^2)/2 + C$
- $\varphi(x, y, z) = x^2/2 - y^2/2 + xz - yz + C$

8. f não é um gradiente
9. f não é um gradiente
10. f não é um gradiente
11. $\varphi(x, y, z) = y^2 \operatorname{sen} x + xz^3 - 4y + 2z + C$
12. $\varphi(x, y, z) = x + 2x^2y - x^3z^2 + 2y - z^3 + C$
13. (b) $\varphi(x, y) = \frac{ar^{n+1}}{n+1} + C \operatorname{sen} n \neq -1$; $\varphi(x, y) = a \log r + C \operatorname{sen} n = -1$
15. $\varphi(x) = \frac{r^{p+2}}{p+2} + C \operatorname{sen} p \neq -2$; $\varphi(x) = \log r + C \operatorname{sen} p = -2$
16. $\varphi(x) = g(r) + C$

10.20. Exercícios (pg. 392)

1. $x^2/2 + 2xy + y^2/2 = C$
2. $x^2y = C$
3. $x^3/3 - xy - y/2 + (\operatorname{sen} 2y)/4 = C$
4. $\cos 2x \operatorname{sen} 3y = C$
5. $x^3y + 4x^2y^2 - 12e^y + 12ye^y = C$
6. $\int Q(x)e^{fP(x)dx} dx - ye^{fP(x)dx} = C$
8. (a) $x + y = Cy^2$
(b) $y^3/x^3 - 3 \log |x| = C$
9. (a) $6(xy)^{1/2} - (y/x)^{3/2} = C$; $(x^5y)^{-1/2}$ é um factor integrante
(b) $x + e^{-x} \operatorname{sen} y = C$; $e^{-x} \cos y$ é um factor integrante
10. $x^3y^4 + x^4y^5 = C$, $10x^3y^4 + x^5y^5 = C$, respectivamente; x^2y^3 é um factor integrante comum

Capítulo 11

11.9. Exercícios (pg. 407)

- | | |
|-------------------------------|--|
| 1. $\frac{1}{3}$ | 7. 6 |
| 2. 1 | 8. $t^{-3}(e^{t^3} - e^t) + t^{-2} - t^{-1}$ |
| 3. $2\sqrt{3} - \frac{32}{3}$ | 10. $\frac{1}{6}$ |
| 4. $\pi^2/4$ | 11. $\frac{1}{6}(\frac{21}{8} - \sqrt{2})$ |
| 5. 2 | 12. $\pi/2$ |
| 6. 2π | 13. $(\log 2)/6$ |

11.15. Exercícios (pg. 417)

- | | |
|--|---|
| 1. $-3\pi/2$ | 5. $\pi^2 - \frac{40}{9}$ |
| 2. $\frac{3}{2} + \cos 1 + \operatorname{sen} 1 - \cos 2 - 2 \operatorname{sen} 2$ | 6. 6 |
| 3. $e - e^{-1}$ | 7. $\frac{80}{3}$ |
| 4. $\frac{7}{2} \log 2$ | 8. (a) $\frac{8}{3}$ (b) 2 (c) 320π |
| 9. $\int_0^1 \left[\int_x^1 f(x, y) dy \right] dx$ | |

10. $\int_0^4 \left[\int_{x/2}^{\sqrt{x}} f(x, y) dy \right] dx$
11. $\int_1^2 \left[\int_1^{y^3} f(x, y) dx \right] dy$
12. $\int_0^1 \left[\int_{2-y}^{1+\sqrt{1-y^2}} f(x, y) dx \right] dy$
13. $\int_{-1}^0 \left[\int_{-\sqrt{4y+4}}^{\sqrt{4y+4}} f(x, y) dx \right] dy + \int_0^8 \left[\int_{-\sqrt{4y+4}}^{2-y} f(x, y) dx \right] dy$
14. $\int_0^1 \left[\int_{e^y}^e f(x, y) dx \right] dy$
15. $\int_{-1}^0 \left[\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} f(x, y) dx \right] dy + \int_0^1 \left[\int_{-\sqrt{1-y}}^{\sqrt{1-y}} f(x, y) dx \right] dy$
16. $\int_0^1 \left[\int_{y^{1/3}}^{y^{1/3}} f(x, y) dx \right] dy$
17. $\int_{-1}^0 \left[\int_{-2\arcsen y}^{\pi} f(x, y) dx \right] dy + \int_0^1 \left[\int_{\arcsen y}^{\pi - \arcsen y} f(x, y) dx \right] dy$
18. $\int_{-2}^0 \left[\int_{2x+4}^{4-x^2} f(x, y) dy \right] dx$
19. $\int_0^1 \left[\int_x^{2-x} (x^2 + y^2) dy \right] dx = \frac{4}{3}$
20. $y = 0, \quad y = x \tan c, \quad x^2 + y^2 = a^2, \quad x^2 + y^2 = b^2$
21. (a) $\int_1^8 \left[\int_{y^{1/3}}^y f(x, y) dx \right] dy$
(b) $4e^8 + 2e/3$
22. $m = 2; \quad n = 1$

11.18. Exercícios (pg. 424)

1. $\bar{x} = -\frac{1}{2}, \quad \bar{y} = \frac{8}{9}$
2. $\bar{x} = 1, \quad \bar{y} = 0$
3. $\bar{x} = \frac{18}{5}, \quad \bar{y} = \frac{50}{9}$
4. $\bar{x} = \pi/2, \quad \bar{y} = \pi/8$
5. $\bar{x} = (\sqrt{2} + 1) \left(\frac{\pi\sqrt{2}}{4} - 1 \right) = \frac{\pi}{2} + \frac{\pi\sqrt{2}}{4} - 1 - \sqrt{2}, \quad \bar{y} = \frac{\sqrt{2} + 1}{4}$
6. $\bar{x} = \frac{2a^2 \log a - a^2 + 1}{4(a \log a - a + 1)}, \quad \bar{y} = \frac{a(\log a)^2}{2(a \log a - a + 1)} - 1$
7. $\bar{x} = \bar{y} = \frac{1}{8}$
8. $\bar{x} = \bar{y} = 256/(315\pi)$
9. $\frac{2a}{3} - \frac{1}{2} \log 3$
10. $\bar{x} = \frac{2}{3} \|\vec{AB}\|, \quad \bar{y} = \frac{2}{3} \|\vec{AD}\|$; supondo que os eixos OX e OY são escolhidos de modo a que coincidam com os lados AB e AD , respectivamente.

11. $I_x = \frac{5\pi}{12}$, $I_y = \frac{2\pi^3}{3} - \pi$
12. $I_x = \frac{1}{12}b^3(a-c)$, $I_y = \frac{1}{12}b(a^3 - c^3)$
13. $I_x = I_y = (1 - 5\pi/16)r^4$
14. $I_x = I_y = \frac{9}{8}$
15. $I_x = \frac{1}{64}[(4a-1)e^{4a} - 1]$, $I_y = \frac{1}{8^2}[(a^3 - 3a^2 + 6a - 6)e^{2a} + 6]$
16. $I_x = \frac{1}{10^2}$, $I_y = \frac{1}{4^2}$
19. $\frac{1}{3}h[\sqrt{2} + \log(1 + \sqrt{2})]$
20. $h^2 + \frac{1}{2}r^2$
21. (a) $(\frac{1}{6}, 1)$
 (b) $(\frac{7}{3}, \frac{9}{2})$
 (c) $(\frac{1}{4}, \frac{1}{4})$
 (d) $(\frac{1}{8}, \frac{1}{8})$
22. $h = 2\sqrt{3}$
23. $h > r\sqrt{2}$

11.22. Exercícios (pg. 433)

1. (a) -4
 (b) 4
 (c) 8
 (d) 4π
 (e) $3\pi/2$
2. 0
3. $n = 3$
4. $-\pi$
9. $g(x, y) = \pm [P^2(x, y) + Q^2(x, y)]^{1/2}$

★ 11.25. Exercícios (pg. 439)

1. (b) 0
2. 0, 2π , -2π
3. No máximo três
4. No máximo sete
5. (a) -3
6. 2π

11.28. Exercícios (pg. 449)

1. $\int_0^{2\pi} \left[\int_0^a f(r \cos \theta, r \sin \theta) r dr \right] d\theta$
2. $\int_{-\pi/2}^{\pi/2} \left[\int_0^{2 \cos \theta} f(r \cos \theta, r \sin \theta) r dr \right] d\theta$
3. $\int_0^{2\pi} \left[\int_a^b f(r \cos \theta, r \sin \theta) r dr \right] d\theta$

4. $\int_0^{\pi/2} \left[\int_0^{g(\theta)} f(r \cos \theta, r \sin \theta) r dr \right] d\theta$, onde $g(\theta) = 1/(\cos \theta + \sin \theta)$
5. $\int_0^{\pi/4} \left[\int_0^{\tan \theta \sec \theta} f(r \cos \theta, r \sin \theta) r dr \right] d\theta + \int_{\pi/4}^{3\pi/4} \left[\int_0^{\csc \theta} f(r \cos \theta, r \sin \theta) r dr \right] d\theta$
 $+ \int_{3\pi/4}^{\pi} \left[\int_0^{\tan \theta \sec \theta} f(r \cos \theta, r \sin \theta) r dr \right] d\theta$
6. $\frac{3}{4}\pi a^4$
7. $\frac{1}{6}a^3[\sqrt{2} + \log(1 + \sqrt{2})]$
8. $\sqrt{2} - 1$
9. $\pi a^4/8$
10. $\int_0^{\pi/4} \left[\int_0^{\sec \theta} f(r \cos \theta, r \sin \theta) r dr \right] d\theta + \int_{\pi/4}^{\pi/2} \left[\int_0^{\csc \theta} f(r \cos \theta, r \sin \theta) r dr \right] d\theta$
11. $\int_{\pi/4}^{3\pi/4} \left[\int_0^{2 \sec \theta} f(r) r dr \right] d\theta$
12. $\int_0^{\pi/2} \left[\int_{g(\theta)}^1 f(r \cos \theta, r \sin \theta) r dr \right] d\theta$, onde $g(\theta) = 1/(\cos \theta + \sin \theta)$
13. $\int_0^{\pi/4} \left[\int_{\tan \theta \sec \theta}^{\sec \theta} f(r \cos \theta, r \sin \theta) r dr \right] d\theta$
14. $\pi^4/3$
15. (a) $u = 7x - y$, $v = -5x + y$
 (b) 60
17. (a) $1 + 2u$
 (c) $\frac{1}{3}$
 (d) $2 + \frac{2}{\sqrt{3}} \left(\arctan \frac{1}{\sqrt{3}} - \arctan \frac{5}{\sqrt{3}} \right)$
18. (a) $4(u^2 + v^2)$
 (c) 0
19. $\frac{\pi}{1-p} [(p^2 + r^2)^{1-p} - p^{2(1-p)}]$ se $p \neq 1$; $\pi \log(1 + r^2)$ se $p = 1$.

$I(p, r)$ tende para um limite finito, quando $p > 1$

11.34. Exercícios (pg. 463)

1. $\frac{1}{364}$
2. $\log \sqrt{2} - \frac{5}{16}$
3. $\frac{1}{48}$
4. $\frac{4}{5}\pi abc$
5. $\pi/6$
6. $\int_0^1 \left\{ \int_0^x \left[\int_0^{1-x} f(x, y, z) dy \right] dz + \int_x^1 \left[\int_{z-x}^{1-x} f(x, y, z) dy \right] dz \right\} dx$
7. $\int_0^1 \left\{ \int_{-x}^x \left[\int_{-\sqrt{z^2-x^2}}^{\sqrt{z^2-x^2}} f(x, y, z) dy \right] dx \right\} dz$

8. $\int_0^1 \left\{ \int_0^{x^2} \left[\int_0^1 f(x, y, z) dy \right] dz + \int_{x^2}^{1+x^2} \int_{\sqrt{z-x^2}}^1 \left[f(x, y, z) dy \right] dz \right\} dx$
10. $16\pi/3$
11. $\frac{1}{6}$
12. $\frac{1}{60}\pi a^2 h(3a^2 + 2h^2)$
13. $\frac{4}{3}\pi a^3$
14. $\frac{4}{3}\pi(b^3 - a^3)$
15. $\frac{2}{3}\pi R^3(a^2 + b^2 + c^2)^{-1/2}$
18. $\frac{2}{3}\pi(5\sqrt{5} - 4)$
19. $\frac{32}{9}$
20. $\frac{4}{5}\pi(b^5 - a^5)$
22. No eixo, a uma distância $\frac{1}{3}h$ da base
23. No eixo, a uma distância $\frac{1}{3}h$ da base
24. No eixo de simetria a uma distância $\frac{3}{8} \cdot \frac{b^4 - a^4}{b^3 - a^3}$ do plano secante dos hemisférios
25. $\bar{x} = \bar{y} = \bar{z} = \frac{7}{12}h$. (supondo o vértice referido na origem)
26. $\frac{3}{20}M(a^2 + 4h^2)$
27. $\frac{2}{5}MR^2$
28. $\frac{3}{5}Ma^2$
29. $2^{1/4}$

Capítulo 12

12.4. Exercícios (pg. 475)

1. $(a_2b_3 - a_3b_2)(x - x_0) + (a_3b_1 - a_1b_3)(y - y_0) + (a_1b_2 - a_2b_1)(z - z_0) = 0$;
 $\frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v} = (a_2b_3 - a_3b_2)\mathbf{i} + (a_3b_1 - a_1b_3)\mathbf{j} + (a_1b_2 - a_2b_1)\mathbf{k}$
2. $x^2/a^2 + y^2/b^2 = z$; $\frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v} = -2bu^2 \cos v \mathbf{i} - 2au^2 \sin v \mathbf{j} + abu \mathbf{k}$
3. $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$; $\frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v} = abc \sin u \left(\frac{\sin u \cos v}{a} \mathbf{i} + \frac{\sin u \sin v}{b} \mathbf{j} + \frac{\cos u}{c} \mathbf{k} \right)$
4. $z = f(\sqrt{x^2 + y^2})$; $\frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v} = -uf'(u) \cos v \mathbf{i} - uf'(u) \sin v \mathbf{j} + uk$
5. $\frac{y^2}{a^2} + \frac{z^2}{b^2} = 1$; $\frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v} = b \sin v \mathbf{j} + a \cos v \mathbf{k}$
6. $(\sqrt{x^2 + y^2} - a)^2 + z^2 = b^2$;
 $\frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v} = b(a + b \cos u)(\cos u \sin v \mathbf{i} + \cos u \cos v \mathbf{j} + \sin u \mathbf{k})$
7. $|abc| \cosh v \left[\left(\frac{\sinh^2 u}{a^2} + \frac{\cosh^2 u}{b^2} \right) \cosh^2 v + \frac{\sinh^2 v}{c^2} \right]^{1/2}$

8. $\sqrt{128v^2 + 4}$
9. $|u - v| \sqrt{36u^2v^2 + 9(u + v)^2 + 4}$
10. $\sqrt{u^4 + u^2}$

12.6. Exercícios (pg. 481)

2. $\pi a^2 \sqrt{3}$
3. $(2\pi - 4)a^2$
4. 4
5. (a) Um parabolóide de revolução
(b) $-2u^2 \cos v\mathbf{i} - 2u^2 \sin v\mathbf{j} + u\mathbf{k}$
(c) $n = 6$
6. $\sqrt{2} \pi a^2/4$
7. $2\pi\sqrt{6}$
8. $2\pi a^2(3\sqrt{3} - 1)/3$
9. $4\pi^2 ab$
11. (a) Um círculo de raio unidade no plano XOY ; um semi-círculo de raio unidade no plano XOZ , com $z \leq 0$; um semi-círculo unitário no plano $x = y$ com $z \leq 0$
(b) A semiesfera $x^2 + y^2 + z^2 = 1, z \leq 0$
(c) A esfera $x^2 + y^2 + z^2 = 1$ excepto o polo norte; a linha que une o polo norte com (x, y, z) intercepta o plano XOY em $(u, v, 0)$

12.10. Exercícios (pg. 489)

1. $4\pi/3$
3. $\bar{x} = \bar{y} = \bar{z} = a/2$
4. $\frac{1}{2}$
7. 0
8. $\pi\sqrt{2}$
9. No eixo do cone, a uma distância $\frac{1}{4}a(1 - \cos \alpha)/(1 - \cos(\alpha/2))$ do centro da esfera
10. $\pi a^3 h + \frac{2}{3}\pi a h^3$
12. $2\pi/3$
11. $3\pi a^3 h + \frac{2}{3}\pi a h^3$
13. $-\pi/3$

12.13. Exercícios (pg. 495)

1. 0
3. -4
2. $-\pi$
4. $\frac{4}{3}$

12.15. Exercícios (pg. 500)

1. (a) $\text{div } F(x, y, z) = 2x + 2y + 2z$; $\text{rot } F(x, y, z) = \mathbf{0}$
(b) $\text{div } F(x, y, z) = 0$; $\text{rot } F(x, y, z) = 2\mathbf{i} + 4\mathbf{j} + 6\mathbf{k}$
(c) $\text{div } F(x, y, z) = -x \sin y$; $\text{rot } F(x, y, z) = \mathbf{i} + \mathbf{j}$

- (d) $\operatorname{div} F(x, y, z) = ye^{xy} - x \operatorname{sen}(xy) - 2xz \operatorname{sen}(xz^2)$;
 $\operatorname{rot} F(x, y, z) = z^2 \operatorname{sen}(xz^2)\mathbf{j} - [xe^{xy} + y \operatorname{sen}(xy)]\mathbf{k}$
- (e) $\operatorname{div} F(x, y, z) = 2x \operatorname{sen} y + 2y \operatorname{sen}(xz) - xy \operatorname{sen} z \cos(\cos z)$;
 $\operatorname{rot} F(x, y, z) = [x \operatorname{sen}(\cos z) - xy^2 \cos(xz)]\mathbf{i} - y \operatorname{sen}(\cos z)\mathbf{j} + [y^2 z \cos(xz) - x^2 \cos y]\mathbf{k}$

2. 0
 4. $n = -3$
 5. Não existe tal campo vectorial
 10. Um tal campo é $\mu(x, y, z) = (xyz)^{-2}$
 11. $\operatorname{div}(V \times r) = 0$; $\operatorname{rot}(V \times r) = (c + 1)V$
 13. $16(a + b)$

★ 12.17. Exercícios (pg. 506)

1. $(3x - 2z)\mathbf{j} - x\mathbf{k}$ é um tal campo
 2. $(x^2/2 - xy - yz + z^2/2)\mathbf{j} + (x^2/2 - xz)\mathbf{k}$ é um tal campo
 3. $(x^2y/2 + z^2/2)\mathbf{j} + \nabla f(x, y)$ para qualquer f independente de z
 5. $G(x, y, z) = \frac{yz}{r(x^2 + y^2)}\mathbf{i} - \frac{xz}{r(x^2 + y^2)}\mathbf{j}$ satisfaz a $\operatorname{rot} G = r^{-1}\mathbf{r}$ em todos os pontos não pertencentes ao eixo OZ
 6. $f(r) = Cr^{-3}$
 9. $F(x, y, z) = -\frac{1}{3}(z^3\mathbf{i} + x^3\mathbf{j} + y^3\mathbf{k})$, $G(x, y, z) = \frac{1}{3}\nabla(x^3y + y^3z + z^3x)$
 10. (c) $3\pi/2$

12.21. Exercícios (pg. 517)

1. 3
 2. (a) 144π
 (b) -16π
 (c) 128π
 15. 8π
3. (a) $3|V|$
 (b) $9|V|\bar{z}$
 (c) $|V|\bar{x}$
 (d) $4I_x$

Capítulo 13

13.4. Exercícios (pg. 529)

2. $A_1 \cup A_2 \cup A_3 = (A_1 \cap A_2' \cap A_3') \cup (A_2 \cap A_3') \cup A_3$; $\bigcup_{k=1}^n A_k = \bigcup_{k=1}^{n-1} (A_k \cap \bigcap_{j=k+1}^n A_j') \cup A_n$
3. (i) (ii) (iii) (iv) (v)

(a)	$A' \cap B'$	$A \cap B'$	$A \cup B$	$(A \cap B') \cup (A' \cap B)$	$A' \cup B'$
(b)	500	200	500	300	800

6. $\mathcal{B}_2 = \{\emptyset, A_1, A_2, A_1 \cup A_2, A'_1, A'_2, A'_1 \cap A'_2, S\}$
 7. $\mathcal{B}_3 = \{\emptyset, A_1, A_2, A_3, A_1 \cup A_2, A_2 \cup A_3, A_1 \cup A_3, A_1 \cup A_2 \cup A_3, A'_1, A'_2, A'_3, A'_1 \cap A'_2, A'_2 \cap A'_3, A'_1 \cap A'_3, A'_1 \cap A'_2 \cap A'_3, S\}$ (if $n > 3$)

13.7. Ejercicios (pg. 534)

- $A \subseteq B'$
- $x \in A' \cap B' \cap C'$
- $x \in A \cap B' \cap C'$
- $x \in A \cup B \cup C$
- $x \in (A \cap B' \cap C') \cup (A' \cap B \cap C') \cup (A' \cap B' \cap C)$
- $x \in (A' \cap B') \cup (B' \cap C') \cup (A' \cap C')$
- $x \in (A \cap B) \cup (A \cap C) \cup (B \cap C)$
- $x \in (A \cap B \cap C') \cup (A \cap B' \cap C) \cup (A' \cap B \cap C)$
- $x \in (A \cap B \cap C)'$
- $x \in A \cap C' \cap B'$
- $x \in A \cap B \cap C$
- $x \in A \cup B \cup C$
- $1 - a$
 - $1 - b$
 - $a + b - c$
 - $1 - c$
 - $1 - a + c$
 - $a - c$

13.9. Ejercicios (pg. 13.9)

- $\frac{5}{14}$
 - $\frac{45}{91}$
 - $\frac{10}{91}$
 - $\frac{36}{91}$
- $\frac{23}{36}$
 - $\frac{1}{6}$
 - $\frac{1}{3}$
 - $\frac{5}{36}$
 - $\frac{13}{36}$
 - $\frac{2}{3}$
- $\frac{1}{4}$
- $P_0 = 1 - P(A) - P(B) + P(A \cap B), P_1 = P(A) + P(B) - 2P(A \cap B), P_2 = P(A \cap B)$
- 5 á 9
 - 45 á 46
 - 10 á 81
 - 36 á 55
- $\frac{9}{47}$
 - $\frac{8}{47}$
 - $\frac{4}{47}$
 - $A/(A+B)$
 - $B/(A+B)$
 - $(C+1)/(C+D+1)$
 - $C/(C+D+1)$
- $\frac{1}{9}$
 - $\frac{4}{9}$
 - $\frac{5}{9}$
 - $\frac{8}{9}$

13.11. Ejercicios (pg. 544)

- $\{(1, 2), (1, 3), (2, 1), (2, 3), (3, 1), (3, 2)\}$
- 1326
- 54

Soluções dos exercícios

4. $\{H, T\} \times \{H, T\} \times \{1, 2, 3, 4, 5, 6\}$; 24 resultados
5. $52!/(13!)^4$
6. 36
 - (a) 18
 - (b) 12
 - (c) 24
7. (a) $13 \cdot 12 \cdot 11 \cdot 72 = 123552$ (não incluindo grupos de três e de quatro)
 - (b) 5148
 - (c) 36 (não incluindo 10 JQKA)
 - (d) 4

8. (a) $4 \binom{13}{5} / \binom{52}{5}$ (b) $36 / \binom{52}{5}$ (c) $4 / \binom{52}{5}$

9. (a) $\frac{2 \cdot 98!}{(49!)^2}$ (b) $\frac{98!}{48! \cdot 50!}$

10. $\binom{98}{48} / \binom{100}{50}$

11. 16

12. n^k

13.14. Exercícios (pg. 549)

2. (a) $P(A) = \frac{3}{10}$; $P(B|A) = \frac{6}{11}$; $P(A \cap B) = \frac{9}{55}$

4. $\frac{\binom{98}{48}}{\binom{100}{50} - \binom{98}{50}}$

5. $\frac{4}{11}$

6. (a) $1 - \frac{26! \cdot 34!}{21! \cdot 39!} = 1 - \frac{\binom{34}{13}}{\binom{39}{13}} = 1 - \frac{\binom{62}{5}}{\binom{39}{5}}$

- (b) $1 - \frac{\binom{34}{13} + 5 \binom{34}{12}}{\binom{39}{13}}$ (c) $\frac{3 \binom{26}{5}}{\binom{39}{5} \binom{39}{13}}$

9. $\frac{1}{2}$

15. (a) $P(A) = P(B) = P(C) = \frac{1}{2}$; $P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{1}{4}$; $P(A \cap B \cap C) = 0$

13.18. Exercícios (pg. 560)

- $P(H, H) = p_1 p_2$; $P(H, T) = p_1(1 - p_2)$; $P(T, H) = (1 - p_1)p_2$; $P(T, T) = (1 - p_1)(1 - p_2)$
 - Sim
 - Não
 - H_1 e H_2 , H_1 e T_2 , H_2 e T_1 , T_1 e T_2
- $\frac{319}{512}$
 - $\frac{11}{1024}$
 - 6
- $\binom{10}{3} \frac{5^7}{6^{10}} = \frac{390625}{2519424}$
- $\frac{5}{16}$
 - $\frac{1}{2}$
 - $\frac{3}{16}$
- $(5!)^2/10! = \frac{1}{252}$
 - $\frac{1}{2}$
- $36p^{10} - 80p^9 + 45p^8$
 - $\frac{7}{128}$
- É vantajoso aportar importâncias iguais
- $f(p) = (1 - p)^2 + p^3$
 - $(\sqrt{31} - 4)/3$
- $\binom{n}{k} \frac{w^k b^{n-k}}{(w+b)^n}$
 - $\binom{8}{3} \frac{17^5}{18^8} = \frac{9938999}{1377495072}$
 - $\frac{103}{512}$
 - $1 - (19/20)^{10} = 0.4013$
 - $\frac{193}{512}$
 - $59 \leq n \leq 75$

13.20. Exercícios (pg. 566)

- $f(k) = 2k$
 - $f(k) = 3^k$
 - $f(k) = p_k$, onde p_k é o k -ésimo número primo ≥ 2
 - uma tal função é $f(k) = (g(k), h(k))$, onde

$$g(k) = \frac{m^2(k) + 3m(k)}{2} - k + 2, \quad h(k) = k - \frac{m^2(k) + m(k)}{2},$$

e

$$m(k) = \left\lfloor \frac{\sqrt{8k - 7} - 1}{2} \right\rfloor$$

onde $\lfloor x \rfloor$ representa o maior inteiro $\leq x$.(e) $f(k) = 2^{g(k)} 3^{h(k)}$, onde $g(k)$ e $h(k)$ são as definidas em (d)

13.22. Exercícios (pg. 569)

- $n = 0$: $\max = 1$, $\min = \frac{1}{2}$
 $n = 1$: $\max = \frac{1}{4}$, $\min = 0$
 $n = 2$: $\max = \frac{1}{8}$, $\min = 0$
 $n = 3$: $\max = \frac{1}{16}$, $\min = 0$

3. (a) $1 - qp^3 - pq^3$
 (b) $\frac{7}{8}$
 (c) $\frac{3}{8}$

4. (a) $3pq/(pq + 2)$
 (b) $\frac{1}{3}$
 (c) $2 \log 2 - 1$

13.23. Exercícios variados sobre probabilidade. (pg. 569)

1. $\frac{1}{12}$
 2. (a) $\frac{2}{19}$
 (b) $\frac{1}{10}$
 3. (a) $\frac{1}{4}$
 (b) $\frac{7}{12}$
 4. (a) $\frac{15}{64}$
 (b) $\frac{13}{51}$
 (c) $\frac{12}{163}$
 5. $\frac{64}{73}$
 7. 0.65

8. (a) $\frac{2}{9}$
 (b) $\frac{2}{9}$
 (c) $\frac{2}{9}$
 (d) Não

9. $p^3 + 6p\left(\frac{1-p}{2}\right)^2$
 10. $np(1-p)^{n-1} + np^{n-1}(1-p)$
 11. $\frac{n}{2^{n-1}} \left(1 - \frac{n}{2^{n-1}}\right)^{n-1}$

Capítulo 14

14.4. Exercícios (pg. 577)

1. (b) $X \leq b$
 2. (a) $\{\omega \mid X(\omega) \in (a, b], Y(\omega) \in (c, d]\}$
 (c) $X \leq a, Y \leq d$
 (d) $P(a < X \leq b, c < Y \leq d) = P(X \leq b, Y \leq d) - P(X \leq a, Y \leq d) - P(X \leq b, Y \leq c) + P(X \leq a, Y \leq c)$
 3. (a) $\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}, \{(5, 6), (6, 5)\}, \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1), (5, 6), (6, 5)\}$
 (b) $P(X = 7) = \frac{1}{8}; P(X = 11) = \frac{1}{8}; P(X = 7 \text{ or } X = 11) = \frac{2}{8}$
 4. $Y = X_1 + X_2 + X_3 + X_4; P(Y = 0) = \frac{1}{16}; P(Y = 1) = \frac{1}{4}; P(Y \leq 1) = \frac{5}{16}$
 5. $Y = 7X$ se $0 \leq X \leq 100; Y = 10X - 300$ se $X > 100$
 6. (a) $Z = Y - 1$
 (b) $U = Y_1 + Y_2 - 1$

14.8. Exercícios (pg. 588)

2. t	2	3	4	5	6	7	8	9	10	11	12
$p_x(t)$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

3. (b) $p(-2) = \frac{1}{2}, p(0) = p(2) = \frac{1}{4}$
 (c) $0, \frac{3}{4}, \frac{3}{4}, \frac{1}{4}, 1, 0, \frac{1}{4}, \frac{1}{4}$
 5. (a) $c = \frac{1}{3}$
 (b) $\frac{1}{9}, \frac{1}{3}, \frac{2}{3}, \frac{8}{9}$
 (c) $F(t) = 0$ para $t < 0, F(t) = \frac{1}{3}$ para $0 \leq t < 1, F(t) = \frac{1}{3}$ para $1 \leq t < 2, F(t) = 1$ para $t \geq 2$

- (d) Não existe tal t
 (e) $t = 2$
6. (a)

k	0	1	2	3	4
$p(k)$	$\frac{1}{81}$	$\frac{2}{81}$	$\frac{8}{27}$	$\frac{8}{81}$	$\frac{1}{81}$

 $p(t) = 0$ para $t \neq 0, 1, 2, 3, 4$
- (b) $F(t) = 0$ para $t < 0$; $F(t) = \frac{1}{81}$ em $[0, 1)$; $F(t) = \frac{1}{27}$ em $[1, 2)$; $F(t) = \frac{8}{81}$ em $[2, 3)$; $F(t) = \frac{5}{81}$ em $[3, 4)$; $F(t) = 1$ para $t \geq 4$
- (c) $\frac{8}{27}, \frac{5}{81}$
7. (b) $P(X = 0) = (1 - p)^2$; $P(X = 1) = 2p(1 - p)$
8. $p_X(k) = \frac{2k}{n(n+1)}$; $F_X(t) = \frac{[t]([t] + 1)}{n(n+1)}$ para $0 \leq t \leq n$, onde $[t]$ é o maior inteiro $\leq t$; $F_X(t) = 0$ para $t < 0$; $F_X(t) = 1$ para $t > n$
9. $p(k) = e^{-c} \frac{c^k}{k!}$; $k = 0, 1, 2, 3, \dots$; $c \geq 0$
 $p(t) = 0$ para $t \neq 0, 1, 2, 3, \dots$
10. (a) $p_X(t) = \frac{1}{2}$ em $t = -1$ e $t = +1$; $p_X(t) = 0$ para os restantes valores
 $F_X(t) = 0$ para $t < -1$; $F_X(t) = \frac{1}{2}$ para $-1 \leq t < 1$; $F_X(t) = 1$ para $t \geq 1$
11. $P(A) = \frac{2}{3}$; $P(B) = \frac{2}{3}$; $P(A \cap B) = \frac{1}{3}$; $P(B|A) = \frac{1}{2}$; $P(A \cup B) = 1$

14.12. Exercícios (pg. 598)

1. (a) $c = 1$; $f(t) = 1$ se $0 \leq t \leq 1$; $f(t) = 0$ para os restantes valores de t .
 (b) $0, \frac{1}{3}, \frac{1}{3}$
2. $c = \frac{1}{2}$; $F(t) = 0$ se $t < -\pi/2$; $F(t) = \frac{1}{2} \cos t$ se $-\pi/2 \leq t < 0$; $F(t) = 1 - \frac{1}{2} \cos t$ se $0 \leq t < \pi/2$; $F(t) = 1$ se $t \geq \pi/2$
3. $c = \frac{3}{4}$; $F(t) = 0$ se $t < 0$; $F(t) = \frac{1}{4}(3t^2 - t^3)$ se $0 \leq t \leq 2$; $F(t) = 1$ se $t > 2$
4. (a) $\frac{1}{2}$
 (b) $\frac{1}{2}$
 (c) $\frac{1}{4}$
5. (a) $f(t) = 0$ se $t < \frac{1}{4}$; $f(t) = 8t - 1$ se $\frac{1}{4} \leq t < \frac{1}{2}$; $f(t) = 7 - 8t$ se $\frac{1}{2} \leq t < \frac{3}{4}$; $f(t) = 0$ se $t \geq \frac{3}{4}$
 (b) $F(t) = 0$ se $t < \frac{1}{4}$; $F(t) = 4t^2 - t$ se $\frac{1}{4} \leq t < \frac{1}{2}$; $F(t) = -4t^2 + 7t - 2$ se $\frac{1}{2} \leq t < \frac{3}{4}$; $F(t) = 1$ se $t \geq \frac{3}{4}$
 (c) $1, 1, \frac{1}{2}, 0, \frac{1}{16}$
6. (a) $0, \frac{5}{8}, \frac{3}{8}, \frac{1}{2}$
 (b) $t = 1$
7. (a)

k	5	10	15	20	25	30
$P(X \geq k)$	$\frac{5}{6}$	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$	0
8. $F_Y(t) = 0$ se $t < b$; $F_Y(t) = (t - b)/a$ se $b \leq t \leq b + a$; $F_Y(t) = 1$ se $t > b + a$
9. (a) $\frac{2}{37}$
 (b) $\frac{11}{37}$
 (c) $\frac{36}{37}$

$$10. F(t) = \frac{1}{2} + \frac{1}{\pi} \arctg \frac{t-b}{a}$$

14.16. Exercícios (pg. 607)

1. (a) 105
(b) 10,05
2. (a) $\frac{1}{2}(\sqrt{2} - 1)$
(b) $\frac{1}{4}(2 - \sqrt{2})$
(c) $\frac{1}{2}(2 - \sqrt{2})$
(d) $\frac{1}{4}(4 - \sqrt{2})$
3. (a) $1 - e^{-1}$
(b) e^{-2}
(c) $(e - 1)/e^2$
(d) e^{-3}
4. $F(t) = 0$ se $t < c$; $F(t) = 1 - e^{-\lambda(t-c)}$ se $t \geq c$
6. $\lambda' = \lambda/a$, $c' = b + ac$
9. (a) 0,5000
(b) 0,1359
(c) 0,9974
(d) 0,0456
10. (a) 0,675
(b) 0,025
11. (a) $0,675 \sigma$
(b) $0,025 \sigma$
12. (a) 0,8185
(b) 0,8400
13. 75,98 polegadas
14. média = b , variância = a^2
15. $F_Y(t) = 0$ se $t < 0$; $f_Y(t) = 0$ se $t \leq 0$; $f_Y(t) = e^{-t/2}/\sqrt{2\pi t}$ se $t > 0$

14.18. Exercícios (pg. 610)

1. (a) $F_Y(t) = 0$ se $t < 1$; $F_Y(t) = (t-1)/3$ se $1 \leq t \leq 4$; $F_Y(t) = 1$ se $t > 4$; $f_Y(t) = \frac{1}{3}$ se $1 \leq t \leq 4$; $f_Y(t) = 0$ para os restantes valores de t .
(b) $F_Y(t) = 0$ se $t < -2$; $F_Y(t) = (t+2)/3$ se $-2 \leq t \leq 1$; $F_Y(t) = 1$ se $t > 1$; $f_Y(t) = \frac{1}{3}$ se $-2 \leq t \leq 1$; $f_Y(t) = 0$ para os restantes valores de t .
(c) $F_Y(t) = 0$ se $t < 0$; $F_Y(t) = t^{1/2}$ se $0 \leq t \leq 1$; $F_Y(t) = 1$ se $t > 1$; $f_Y(t) = (2t)^{-1/2}$ se $0 \leq t \leq 1$; $f_Y(t) = 0$ para os restantes valores de t .
(d) $F_Y(t) = e^t$ se $t \leq 0$; $F_Y(t) = 1$ se $t > 0$; $f_Y(t) = e^t$ se $t \leq 0$; $f_Y(t) = 0$ se $t > 0$
(e) $F_Y(t) = e^{t/2}$ se $t \leq 0$; $F_Y(t) = 1$ se $t > 0$; $f_Y(t) = \frac{1}{2}e^{t/2}$ se $t \leq 0$; $f_Y(t) = 0$ se $t > 0$
(f) $F_Y(t) = 0$ se $t < 1$; $F_Y(t) = \log t$ se $1 \leq t \leq e$; $F_Y(t) = 1$ se $t > e$; $f_Y(t) = 1/t$ se $1 \leq t \leq e$; $f_Y(t) = 0$ para os restantes valores de t .
2. Seja ψ a função inversa de ϕ definida no intervalo aberto (a, b) .
Então $F_Y(t) = 0$ se $t \leq a$; $F_Y(t) = F_X[\psi(t)]$ se $a < t < b$; $F_Y(t) = 1$ se $t \geq b$; $f_Y(t) = f_X[\psi(t)]\psi'(t)$ se $a < t < b$; $f_Y(t) = 0$ para os restantes valores de t .

3. (a) $f_Y(t) = 0$ se $t \leq 0$; $f_Y(t) = (2\pi t)^{-1/2} e^{-t/2}$ se $t > 0$
 (b) $f_Y(t) = 0$ se $t < 0$; $f_Y(t) = 4t(2\pi)^{-1/2} e^{-t^2/2}$ se $t \geq 0$
 (c) $f_Y(t) = 0$ se $t \leq 0$; $f_Y(t) = (2\pi t^2)^{-1/2} e^{-(\log t)^2/2}$ se $t > 0$
 (d) $f_Y(t) = (2\pi)^{-1/2} \sec^2 t e^{-(\tan^2 t)/2}$ se $|t| < \pi/2$; $f_Y(t) = 0$ se $|t| \geq \pi/2$

14.22. Exercícios (pg. 616)

2. (a) $P(X = x_1) = P(X = x_2) = P(Y = y_1) = P(Y = y_2) = \frac{1}{2}(p + q)$
 (b) $p = q = \frac{1}{2}$
3. (a) $F(x, y) = \left(\frac{x-a}{b-a}\right)\left(\frac{y-c}{d-c}\right)$ se $a \leq x \leq b$ e $c \leq y \leq d$,
 $F(x, y) = \frac{x-a}{b-a}$ se $a \leq x \leq b$ e $y > d$, $F(x, y) = \frac{y-c}{d-c}$ se $x > b$ e $c \leq y \leq d$,
 $F(x, y) = 1$ se $x > b$ e $y > d$, $F(x, y) = 0$ para os restantes valores de x e y
 (b) $F_X(x) = (x-a)/(b-a)$ se $a \leq x \leq b$; $F_X(x) = 0$ se $x < a$; $F_X(x) = 1$ se $x > b$;
 $F_Y(y) = (y-c)/(d-c)$ se $c \leq y \leq d$; $F_Y(y) = 0$ se $y < c$; $F_Y(y) = 1$ se $y > d$
 (c) X e Y são independentes
5. $P(Y = 0) = \frac{1}{2}$; $P(Y = 1) = P(Y = 2) = \frac{1}{4}$
7. $\frac{1}{5}, \frac{4}{5}$
8. $\frac{1}{3}$
9. (b) $f(x, y) = \frac{1}{2}$ se $(x, y) \in Q$; $f(x, y) = 0$ se $(x, y) \notin Q$;
 $f_X(x) = 1 - |x|$ se $|x| \leq 1$; $f_X(x) = 0$ se $|x| > 1$;
 $f_Y(y) = 1 - |y|$ se $|y| \leq 1$; $f_Y(y) = 0$ se $|y| > 1$.
 X e Y não são independentes
10. $g(u, v) = f(u + a, v + b)$

14.24. Exercícios (pg. 622)

1. (b) $f_V(v) = 1 + v$ se $-1 \leq v < 0$; $f_V(v) = 1 - v$ se $0 \leq v \leq 1$; $f_V(v) = 0$ se $|v| > 1$
 (c) U e V não são independentes
2. (b) $f_V(t) = 2 - 2t$ se $0 \leq t \leq 1$; $f_V(t) = 0$ para os restantes valores de t
 (c) U e V são independentes
3. (b) $g(u, v) = ue^{-u}$ se $u > 0$, $0 < v < 1$; $g(u, v) = 0$ para os restantes valores de u e v
 (c) $f_U(u) = ue^{-u}$ se $u > 0$; $f_U(u) = 0$ se $u \leq 0$
 (d) $f_V(v) = 1$ se $0 < v < 1$; $f_V(v) = 0$ para os restantes valores de v
4. (b) $g(u, v) = (v/2\pi)e^{-(1+u^2)v^2/2}$ se $v \geq 0$
 (c) $f_U(u) = [\pi(1 + u^2)]^{-1}$
5. $f_Z(t) = \pi^{-1/2} e^{-t^2}$
6. (b) $f_U(u) = 0$ se $u < 0$; $f_U(u) = (u/\sigma^2) \exp\left(-\frac{u^2}{2\sigma^2}\right)$ se $u > 0$; $F_U(t) = 0$ se $t < 0$;
 $F_U(t) = 1 - \exp\left(-\frac{t^2}{2\sigma^2}\right)$ se $t \geq 0$

14.27. Exercícios (pg. 630)

1. $E(X) = \frac{7}{2}$, $\text{Var}(X) = \frac{35}{12}$

7. (a) $E(X) = \text{Var}(X) = \lambda$
 (b) Não
 (c) $E(X) = 1/\lambda$, $\text{Var}(X) = 1/\lambda^2$
 (d) $E(X) = m$, $\text{Var}(X) = \sigma^2$
8. (a) $C(r) = (r-1)/2$
 (b) $F_X(t) = \frac{1}{2}|t|^{1-r} \text{ se } t < -1$; $F_X(t) = \frac{1}{2} \text{ se } -1 \leq t \leq 1$; $F_X(t) = 1 - \frac{1}{2}t^{1-r} \text{ se } t > 1$
 (c) $P(X < 5) = 1 - 5^{r-1}/2$; $P(5 < X < 10) = (5^{1-r} - 10^{1-r})/2$
 (d) X tem esperança matemática finita quando $r > 2$; $E(X) = 0$
 (e) Variâncias finitas para $r > 3$; $\text{var}(X) = (r-1)/(r-3)$
9. $E(X) = E(Y) = -\frac{1}{37}$; $E(Z) = -1767/50653$; $E(X + Y + Z) = -4505/50653$
10. $E(X) \rightarrow \infty$ quando $n \rightarrow \infty$
12. (a) $(2/\pi)^{1/4}$
 (b) $e^{1/4}$
 (c) $e^2 - e$
 (d) $(\pi/2)^{1/4}$

14.31. Exercícios (pg. 639)

4. 251
 5. 0
 6. A desigualdade de Tchebycheff dá $\frac{1}{3}$; as tábuas dão 0,0027
 8. (b) 0,6826
 9. (b) 0,0796
 10. (a) 0,0090
 (b) 0,0179

Capítulo 15

15.5. Exercícios (pg. 650)

2. (a) Não
 (b) Sim
 (c) Sim
 (d) Não
 (e) Sim
 (f) Não
 (g) Sim
3. (a) Nem um nem outro
 (b) Seminorma
 (c) Seminorma
 (d) Nem um nem outro
 (e) Seminorma
 (f) Norma
 (g) Nem um nem outro
 (h) Nem um nem outro
6. (a), (b), (c)
8. (b) O polinômio de (a) mais $\frac{7}{8}(5x^3 - 3x) \int_{-1}^1 (5t^3 - 3t)f(t) dt + \frac{9}{128}(35x^4 - 30x^2 + 3) \times \int_{-1}^1 (35t^4 - 30t^2 + 3)f(t) dt$
9. $a = -3e/4 + 33/(4e)$, $b = 3/e$, $c = \frac{1}{4}e(e - 7/e)$
10. $\frac{1}{128}(1 + 14x^2 - 7x^4)$
11. (b) $\|f - P\|^2 = \frac{1}{60}$
12. (a) $\|P - f\|^2 = \frac{n-1}{n} - \frac{\log^2 n}{n-1}$
 (b) $P(x) = \left(\frac{12}{(n-1)^2} - \frac{6(n+1)}{(n-1)^3} \log n \right) x + \frac{4(n^3-1) \log n}{(n-1)^4} - \frac{6(n+1)}{(n-1)^2}$

$$\|P - f\|^2 = 36 \log 2 - 28 \log^2 2 - \frac{2^3}{2} = 0,0007 \text{ quando } n = 2$$

$$13. (a) \|P - f\|^2 = \frac{1}{2n} [(n-2)e^{2n} + 4e^n - n - 2]$$

$$(b) P(x) = (18 - 6e)x + 4e - 10; \|P - f\|^2 = 20e - \frac{7}{2}e^2 - \frac{57}{2} = 0,0038$$

$$14. (b) \varphi_{k+1}(x) = \frac{\sqrt{(2k+1)(2k+3)}}{k+1} x \varphi_k(x) - \frac{k}{k+1} \sqrt{\frac{2k+3}{2k-1}} \varphi_{k-1}(x), \text{ quando } \varphi_k = P_k / \|P_k\|$$

$$15. (b) \sum_{k=0}^m P_k^2(x) = \frac{P_m}{P_{m+1}} [P_m(x)P'_{m+1}(x) - P_{m+1}(x)P'_m(x)]$$

15.9. Exercícios (pg. 659)

- $P(x) = \frac{1}{8}(x^2 + 13x + 12)$
 - $P(x) = \frac{1}{2}(x^2 - 5x + 6)$
 - $P(x) = -\frac{1}{6}(x^3 - 6x^2 + 5x - 6)$
 - $P(x) = 2x^3 + x^2 - x - 2$
 - $P(x) = -5x^3 - x^2 + 10x - 5$
- $P(x) = \frac{1}{840}(9x^4 - 196x^2 + 640)$
- $Q(x) = 2x^3 + 3x^2 - x - 3$
 - $Q(x) = 4x^3 + 7x^2 - 3x - 7$
- $P(32) = \frac{31}{21}; f(32) - P(32) = \frac{43}{42}$
 - $P(32) = \frac{892}{255}; f(32) - P(32) = -\frac{709}{510}$
 - $P(32) = \frac{43}{15}; f(32) - P(32) = -\frac{11}{30}$
 - $P(32) = -\frac{403}{1630}; f(32) - P(32) = \frac{2114}{765}$
- $L_0(x) = \frac{1}{72}(u-1)(u-3)(u-4)(u-6); L_1(x) = -\frac{1}{30}u(u-3)(u-4)(u-6);$
 $L_2(x) = \frac{1}{18}u(u-1)(u-4)(u-6); L_3(x) = -\frac{1}{24}u(u-1)(u-3)(u-6);$
 $L_4(x) = \frac{1}{180}u(u-1)(u-3)(u-4)$
 - $P(2.6) = 20$
- $x \geq 1,581$
 - $h \leq 0,0006$
- $a = 0, b = 1$
 - $c = 1, d = -2L'_k(x_k)$
- Sejam $B_0(x) = 1$ e $B_n(x) = (x - x_0)(x - x_0 - nh)^{n-1}/n!$ para $n \geq 1$; o único polinômio P de grau $\leq n$ que satisfaz às condições $P(x_0) = c_0, P'(x_1) = c_1, P''(x_2) = c_2, \dots, P^{(n)}(x_n) = c_n$ vem dado por $P(x) = c_0 B_0(x) + \dots + c_n B_n(x)$
- x

15.13. Exercícios (pg. 667)

4. (b)	8	-5040	13068	-13132	6769	-1960	322	-28	1		
	9	40320	-109584	118124	-67284	22449	-4536	546	-36	1	
	10	-3628800	1026576	-1172700	723680	-269325	63273	-9450	870	-45	1

(c) $1 + 2x + 2x^2 - 3x^3 + x^4$

5. (c)

8	1	127	966	1701	1050	266	28	1		
9	1	255	3025	7770	6951	2646	462	36	1	
10	1	511	9330	34105	42525	22827	5880	750	45	1

(d) $-1 + 6x^{(1)} + 16x^{(2)} + 9x^{(3)} + x^{(4)}$

7. (a) $\frac{4}{3}n^3 + \frac{11}{2}n^2 + \frac{61}{6}n$
 (b) $\frac{1}{4}n^4 + \frac{5}{6}n^3 + \frac{3}{4}n^2 + \frac{1}{6}n$
 (c) $\frac{1}{4}n^4 + \frac{5}{6}n^3 + \frac{11}{4}n^2 + \frac{3}{2}n$
 (d) $\frac{1}{5}n^5 + \frac{1}{2}n^4 + \frac{1}{3}n^3 - \frac{1}{30}n$

15.18. Exercícios (pg. 675)

2. (b) $T_n(1) = n^2$, $T'_n(-1) = (-1)^{n-1}n^2$

5. $\sin \theta \sinh \theta = \frac{1-x^2}{n} T'_n(x)$; grau = $n+1$

7. $Q(x) = x^{n+1} - 2^{-n} T_{n+1}(x)$

8. $Q(x) = -\frac{5}{2}x^4 + \frac{3}{16}x^3 - \frac{25}{32}x^2 + \frac{25}{64}x - \frac{1}{512}$

14. (b) $U_0(x) = 1$, $U_1(x) = 2x$, $U_2(x) = 4x^2 - 1$, $U_3(x) = 8x^3 - 4x$,
 $U_4(x) = 16x^4 - 12x^2 + 1$, $U_5(x) = 32x^5 - 32x^3 + 6x$

15.21. Exercícios (pg. 685)

1. (a) $0,693773 - \epsilon$, onde $0,000208 \leq \epsilon \leq 0,001667$. Daqui resultam as desigualdades $0,6921 < \log 2 < 0,6936$
 (b) $n = 578$

2. (a) $c = \sqrt{3}/3$
 (b) $c_1 = \frac{a+b}{2} + \frac{b-a}{2} \frac{\sqrt{3}}{3}$, $c_2 = \frac{a+b}{2} - \frac{b-a}{2} \frac{\sqrt{3}}{3}$

3. (a) $c = \sqrt{2}/2$
 (b) $c_1 = \frac{a+b}{2} + \frac{b-a}{2} \frac{\sqrt{2}}{2}$, $c_2 = \frac{a+b}{2} - \frac{b-a}{2} \frac{\sqrt{2}}{2}$

4. $a = 2 + \sqrt{2}$, $b = 2 - \sqrt{2}$

5. $c = \sqrt{\frac{2}{3}}$

7. (d) $\int_a^b f(x) dx = \frac{b-a}{18} \left[5f\left(\frac{b+a}{2} - \frac{b-a}{2} \sqrt{\frac{3}{5}}\right) + 8f\left(\frac{b+a}{2}\right) + 5f\left(\frac{b+a}{2} + \frac{b-a}{2} \sqrt{\frac{3}{5}}\right) \right]$

10. (a) $\log 2 = 0,693254 - \epsilon$, onde $0,000016 \leq \epsilon \leq 0,000521$; isto dá-nos as desigualdades $0,69273 < \log 2 < 0,69324$
- (b) $\log 2 = 0,69315023 - \epsilon$, onde $0,00000041 \leq \epsilon \leq 0,00001334$; isto conduz-nos às desigualdades $0,6931364 < \log 2 < 0,693149$
11. (d) $\log 2 = 0,693750 - \epsilon$, onde $0,000115 \leq \epsilon \leq 0,003704$; isto conduz-nos às desigualdades $0,69004 < \log 2 < 0,69364$

ÍNDICE ALFABÉTICO

A

- Acontecimento, 533
- Adams, John Couch, 643
- Algebra Boole conjuntos, 527
- operadores coeficientes constantes, 168
- Análise combinatoria, princípios básicos, 539
- erro interpolação polinomial, 656
- numérica, introdução histórica, 643
- Aplicação inversa, 442
- Aplicações biunívocas, 442
- Aproximação elementos espaço euclidiano, 32
- funções contínuas, 33
- polinomial, 644
- — e espaços lineares normados, 646
- —, problemas fundamentais, 648
- Aproximações sucessivas, 251
- — a sistemas não lineares primeira ordem, 255
- — operadores, 261
- Área semi-esfera, 478
- superfície representação paramétrica, 475
- Arquimedes, 644

B

- Bayes, fórmulas (ex. 6), 617
- Bernoulli, distribuição, lei dos grandes números, 635
- , esquema, 555

- Bernoulli, esquema, número mais favorável ocorrências, 557
- , J., 525
- Bessel, equação, 206
- Bolas abertas, 274
- Boole, álgebra conjuntos, 527
- Borel, conjuntos, 574
- , Emile, 574

C

- Cálculo diferencial, aplicações, 319
- — em campos escalares e vectoriais, 273
- e^{it} , 226
- —, casos particulares, 235
- —, método Putzer, 231
- Campo escalar, 273
- —, conjuntos nível, 298
- — continuamente diferenciável, 293
- — diferenciável, 290
- —, gradiente, 291
- —, planos tangentes, 298
- solenoidal, 504
- vectores, condições seja gradiente, 381
- vectorial, 273
- — bidimensional, condição seja gradiente, 430
- —, condições seja gradiente, 382
- —, continuidade e componentes, 279
- —, divergência, 493
- — — nula, 498
- —, reconstrução partindo rotacional, 502

Campo vectorial, rotacional, 493
 — —, — nulo, 498
 — — — solenoidal não rotacional, 504
 Campos escalares contínuos, teorema
 continuidade, 361
 — — —, teorema limitação, 358
 — — —, — valor extremo, 358
 — —, derivação, 307
 — —, fórmula Taylor segunda ordem, 346,
 347
 — —, teorema valores extremos, 360
 — vectoriais, derivadas, 303
 — — —, derivação função composta, 305
 — — e escalares, 273
 Casos favoráveis, 538
 Cauchy, distribuição, 580, 597
 —, solução equação funcional, 594
 Cayley-Hamilton, teorema, 228
 Centro massa, 482
 Complementos ortogonais, 30
 Condição Lindeberg, 639
 — Lipschitz, 257
 Cone, representação paramétrica, 470
 Conjunto fechado, 277
 — finito, 563
 — fundamental, 530
 — infinito numerável, 563
 — limitado medida nula, definição, 409
 — mensurável, 573
 — nível, 299
 — plano simplesmente conexo, definição,
 431
 — sólido, 519
 Conjuntos abertos, 274
 — Borel, 574
 — conexos abertos, 374
 — não numeráveis, 562
 — nível campo escalar, 298
 — numeráveis, 562
 — ortogonais, construção, 25
 — pontos probabilidade positiva, numera-
 bilidade, 574
 Continuidade campos escalares e vectoriais,
 278
 Contradomínio, 37
 Cramer, regra, 105
 Curva, 363
 — nível, 299
 Curvas Jordan, 426

D

D'Alembert, solução equação ondas, 325
 Densidade probabilidade, 614

Derivação campos escalares, 307
 — composição, forma matricial regra, 306
 — função composta, aplicações geométri-
 cas, 298
 — — — para derivadas campos vectoriais,
 305
 — funções compostas para derivadas
 campos escalares, 296
 Derivada campo escalar relativamente um
 vector 284
 — direcciona longo uma curva, 297
 — transformação linear, 285
 Derivadas campos escalares, teorema média,
 285
 — — vectoriais, 303
 — direccionais, 286
 — — e continuidade, 288
 — funções implícitas, 331
 — — parciais, 286
 — —, equações, 319
 — — ordem superior, 287
 — segunda ordem, critério extremos fun-
 ções duas variáveis, 351
 Desigualdade Tchebycheff, 632
 — Wallis, 692
 Desvio padrão, 626
 — quadrático médio, 626
 — tipo, 626
 Determinante, existência função, 100
 — Jacobiano, 442
 — matriz diagonal por blocos, 93
 — — inversa matriz não singular, 92
 — — transposta, 102
 Determinantes, 79
 —, cálculo, 84
 — e independência vectores, 93
 —, fórmulas desenvolvimento, 96
 — jacobianos, 335
 —, menores e complementos algébricos, 96
 —, producto, 91
 —, teorema unicidade, 88
 Diagonalização matriz hermitica ou
 hemi-hermitica, 138
 Diferenciabilidade, condição suficiente,
 293
 — — — continuidade, 304
 Diferencial, 290
 — total, 289
 Descontinuidades funções repartição, 582
 Distribuição Bernoulli, leis dos grandes
 números, 635
 — Cauchy, 580, 597
 — funções duas variáveis aleatórias,
 distribuição, 618
 — normal, 604

- Distribuição normal bivariada (ex. 9), 624
 - Poisson, 588
 - standard, 604
 - uniforme, 580
 - — num intervalo, 592
 - variáveis aleatórias bidimensionais, 610
- Distribuições bidimensionais contínuas, 614
 - contínuas, 591
 - discretas, 585
 - — bidimensionais, 613
 - exponenciais, 599
 - —, soma, 621
 - funções variáveis aleatórias, 608
 - gerais, indicações, 606
 - normais, 602
- Divergência campo vectorial, 493
 - gradiente, 497
 - nula, campo vectorial, 498
 - , propriedades, 496
 - , teorema, 511
 - , —, aplicações, 515
 - um rotacional, 498

E

- Energia potencial, 377
- Equação Bessel, 206
 - característica, 170
 - derivadas parciais primeira ordem com coeficientes constantes, 320
 - — e^{ix} , 222
 - — matricial, teorema unicidade, 223
 - do calor (ex. 1), 329
 - funcional Cauchy, solução, 594
 - indicial, 205
 - Laplace, 319
 - — (ex. 2), 329
 - — n -dimensional (ex. 3), 329
 - Legendre, 194
 - linear homogênea, dimensão espaço solução, 167
 - não homogênea, solução método variação constantes, 178
 - — —, particular método anulador, 185
 - ondas duas dimensões (ex. 7), 331
 - —, solução d'Alembert, 325
 - unidimensional ondas, 324
 - vectorial, 442
- Equações diferenciais lineares, introdução histórica, 161
 - — — ordem n , 165
 - — —, revisão resultados, 162
 - — primeira ordem, aplicações, 389
 - homogêneas e não homogêneas, relação, 177

- Equações lineares coeficientes constantes, soluções por factorização operadores, 170
 - — ordem n , teorema existência e unicidade, 166
 - — segunda ordem com coeficientes analíticos, 191
 - —, sistemas, 66
 - não homogêneas, soluções particulares, 184
- Erro interpolação linear, 658
- Escalar, 5
- Esfera, representação paramétrica, 468
- Espaço amostra, 530
 - euclidiano, norma, 19
 - —, ortogonalidade, 20
 - linear, basas, 13
 - —, componentes, 15
 - —, conjuntos dependentes e independentes, 10
 - —, definição, 3
 - —, dimensão, 13
 - nulo, 37
 - vectorial, 5
- Espaços euclidianos, 16
 - funcionais, 5
 - lineares, 3
 - —, exemplos, 5
 - — matrizes, 58
 - — normados, 262
- Esperança matemática, 625
 - — função variável aleatória, 629
- Esquema Bernoulli, 555
 - —, número mais favorável ocorrências, 557
- Euler, fórmula somação, 688
 - , teorema funções homogêneas, 324
- Extremos condicionados, 353

F

- Fermat, Pierre de, 525
- Fluxo fluido através superfície, 483
- Forma matricial regra derivação composição, 306
 - quadrática à forma diagonal, 145
 - — — — —, aplicações à geometria analítica, 147
- Formas quadráticas, 143
- Fórmula somação Euler, 688
 - Stirling, 692
 - Taylor primeira ordem, 290
 - — segunda ordem campos escalares, 346, 347
- Fréchet, M., 290

Campo vectorial, rotacional, 493
 — —, — nulo, 498
 — — — solenoidal não rotacional, 504
 Campos escalares contínuos, teorema
 continuidade, 361
 — — —, teorema limitação, 358
 — — —, — valor extremo, 358
 — — —, derivação, 307
 — — —, fórmula Taylor segunda ordem, 346,
 347
 — — —, teorema valores extremos, 360
 — vectoriais, derivadas, 303
 — — —, —, derivação função composta, 305
 — — — e escalares, 273
 Casos favoráveis, 538
 Cauchy, distribuição, 580, 597
 —, solução equação funcional, 594
 Cayley-Hamilton, teorema, 228
 Centro massa, 482
 Complementos ortogonais, 30
 Condição Lindeberg, 639
 — Lipschitz, 257
 Cone, representação paramétrica, 470
 Conjunto fechado, 277
 — finito, 563
 — fundamental, 530
 — infinito numerável, 563
 — limitado medida nula, definição, 409
 — mensurável, 573
 — nível, 299
 — plano simplesmente conexo, definição,
 431
 — sólido, 519
 Conjuntos abertos, 274
 — Borel, 574
 — conexos abertos, 374
 — não numeráveis, 562
 — nível campo escalar, 298
 — numeráveis, 562
 — ortogonais, construção, 25
 — pontos probabilidade positiva, numera-
 bilidade, 574
 Continuidade campos escalares e vectoriais,
 278
 Contradomínio, 37
 Cramer, regra, 105
 Curva, 363
 — nível, 299
 Curvas Jordan, 426

D

D'Alembert, solução equação ondas, 325
 Densidade probabilidade, 614

Derivação campos escalares, 307
 — composição, forma matricial regra, 306
 — função composta, aplicações geométricas,
 298
 — — — para derivadas campos vectoriais,
 305
 — funções compostas para derivadas
 campos escalares, 296
 Derivada campo escalar relativamente um
 vector 284
 — direcciona longo uma curva, 297
 — transformação linear, 285
 Derivadas campos escalares, teorema média,
 285
 — — vectoriais, 303
 — direccionais, 286
 — — e continuidade, 288
 — funções implícitas, 331
 — parciais, 286
 — —, equações, 319
 — — ordem superior, 287
 — segunda ordem, critério extremos fun-
 ções duas variáveis, 351
 Desigualdade Tchebycheff, 632
 — Wallis, 692
 Desvio padrão, 626
 — quadrático médio, 626
 — tipo, 626
 Determinante, existência função, 100
 — Jacobiano, 442
 — matriz diagonal por blocos, 93
 — — inversa matriz não singular, 92
 — — transposta, 102
 Determinantes, 79
 —, cálculo, 84
 — e independência vectores, 93
 —, fórmulas desenvolvimento, 96
 — jacobianos, 335
 —, menores e complementos algébricos, 96
 —, producto, 91
 —, teorema unicidade, 88
 Diagonalização matriz hermitica ou
 hemi-hermitica, 138
 Diferenciabilidade, condição suficiente,
 293
 — — continuidade, 304
 Diferencial, 290
 — total, 289
 Discontinuidades funções repartição, 582
 Distribuição Bernoulli, leis dos grandes
 números, 635
 — Cauchy, 580, 597
 — funções duas variáveis aleatórias,
 distribuição, 618
 — normal, 604

Distribuição normal bivariada (ex. 9), 624

- Poisson, 588
 - standard, 604
 - uniforme, 580
 - — num intervalo, 592
 - variáveis aleatórias bidimensionais, 610
- Distribuições bidimensionais contínuas, 614
- contínuas, 591
 - discretas, 585

— — bidimensionais, 613

— exponenciais, 599

— —, soma, 621

— funções variáveis aleatórias, 608

— gerais, indicações, 606

— normais, 602

Divergência campo vectorial, 493

— gradiente, 497

— nula, campo vectorial, 498

—, propriedades, 496

—, teorema, 511

— —, aplicações, 515

— um rotacional, 498

E

Energia potencial, 377

Equação Bessel, 206

— característica, 170

— derivadas parciais primeira ordem com coeficientes constantes, 320

— — e^A , 222

— — matricial, teorema unicidade, 223

— do calor (ex. 1), 329

— funcional Cauchy, solução, 594

— indicial, 205

— Laplace, 319

— — (ex. 2), 329

— — n -dimensional (ex. 3), 329

— Legendre, 194

— linear homogênea, dimensão espaço solução, 167

— não homogênea, solução método variação constantes, 178

— — —, particular método anulador, 185

— ondas duas dimensões (ex. 7), 331

— —, solução d'Alembert, 325

— unidimensional ondas, 324

— vectorial, 442

Equações diferenciais lineares, introdução histórica, 161

— — — ordem n , 165

— — —, revisão resultados, 162

— — primeira ordem, aplicações, 389

— homogêneas e não homogêneas, relação, 177

Equações lineares coeficientes constantes, soluções por factorização operadores, 170

— — ordem n , teorema existência e unicidade, 166

— — segunda ordem com coeficientes analíticos, 191

— —, sistemas, 66

— não homogêneas, soluções particulares, 184

Erro interpolação linear, 658

Escalar, 5

Esfera, representação paramétrica, 468

Espaço amostra, 530

— euclidiano, norma, 19

— —, ortogonalidade, 20

— linear, bases, 13

— —, componentes, 15

— —, conjuntos dependentes e independentes, 10

— —, definição, 3

— —, dimensão, 13

— nulo, 37

— vectorial, 5

Espaços euclidianos, 16

— funcionais, 5

— lineares, 3

— —, exemplos, 5

— — matrizes, 58

— — normados, 262

Esperança matemática, 625

— — função variável aleatória, 629

Esquema Bernoulli, 555

— —, número mais favorável ocorrências, 557

Euler, fórmula somação, 688

—, teorema funções homogêneas, 324

Extremos condicionados, 353

F

Fermat, Pierre de, 525

Fluxo fluido através superfície, 483

Forma matricial regra derivação composição, 306

— quadrática à forma diagonal, 145

— — — —, aplicações à geometria analítica, 147

Formas quadráticas, 143

Fórmula somação Euler, 688

— Stirling, 692

— Taylor primeira ordem, 290

— — segunda ordem campos escalares, 346, 347

Fréchet, M., 290

Frobenius, método, 204
 Função composta, aplicação geométrica
 derivação, 298
 — conjunto finitamente aditiva, 526
 — contínua num conjunto, 278
 — determinante, axiomas, 80, 82
 — —, existência, 100
 — f contínua em a , 278
 — identidade, continuidade, 280
 — potencial, 376
 — —, construção utilizando integrais definidos, 385
 — variável aleatória esperança matemática, 629
 Funções conjunto, 525
 — — completamente aditivas, 526
 — contínuas, integrabilidade, 408
 — densidade, 591, 614
 — distribuição, 579
 — duas variáveis, critério derivadas segunda ordem, 351
 — — — aleatórias, distribuição, 618
 — escada, 398
 — homogêneas, teorema Euler, 324
 — implícitas, derivadas, 331
 — limitadas com descontinuidades, integrabilidade, 409
 — massa probabilística, 585
 — matriciais, conceitos cálculo, 218
 — potenciais conjuntos convexos, 393
 — —, construção, métodos especiais, 384
 — racionais, continuidade, 281
 — repartição, 578
 — —, descontinuidades, 582
 — R^n em R^m , 273
 — variáveis aleatórias, distribuições, 608

G

Gauss-Jordan, determinantes, 87
 — —, método eliminação, 69
 —, teorema, 511
 Gradiente campo escalar, 291
 —, divergência, 497
 —, rotacional, 497
 Gráfico, 363
 Gram-Schmidt, ortogonalização, 25
 Green, George, 425
 —, teorema no plano, 425

H

Herschel, William, 643

Hesse, Ludwig Otto, 346
 Huygens, Christian, 525

I

Igualdade derivadas parciais mistas, condições suficientes, 311
 Imagem, 35
 Independência aleatória, 547
 — da linha, 374
 Integrabilidade funções contínuas, 408
 — — limitadas com descontinuidades, 409
 Integração aproximada, regra trapezoidal, 677
 Integrais duplos, aplicações, 419
 — —, — áreas e volumes, 414
 — —, áreas e volumes, exemplos resolvidos, 415
 — — e inferior, 402
 — — regiões mais gerais, 410
 — linha, 363
 — —, aplicações, 371
 — —, cálculo, primeiro teorema fundamental, 379
 — —, —, segundo teorema fundamental, 374
 — —, introdução, 363
 — —, notações, 364
 — —, propriedades fundamentais, 366
 — — relativos comprimento arco, 370
 — múltiplos, introdução, 397
 — superfície, 467, 481
 — —, mudança representação paramétrica, 484
 — —, notações, 486
 — — triplos, exemplos resolvidos, 459
 Integral duplo, 397
 — —, cálculo, integração unidimensional repetida, 403
 — —, exemplos resolvidos, 405
 — — função definida e limitada retângulo, definição, 401
 — — — escada, 399
 — —, interpretação como volume, 404
 — —, mudança variáveis, 441
 — fluxo, 371
 — inferior, 402
 — Lebesgue-Stieltjes, 607
 — linha, definição, 364
 — —, mudança parâmetro, comportamento, 367
 — —, número giros, 437
 — múltiplo, extensões número superior dimensões, 455

Integral Riemann-Stieltjes, 607
 — superior, 402
 Interpolação, fórmula Newton, 662
 — Lagrange, invariância coeficientes (ex. 6), 659
 — linear, erro, 658
 — polinomial, análise erro, 656
 — —, aplicação fórmula erro, 674
 —, pontos igualmente espaçados, 655, 664
 —, semi-norma, 647
 Invariância integral linha face deformação linha integração, 436
 Inversas, funções, 43
 Isometria, 145
 Isomorfismo transformações lineares matrizes, 59

J

Jordan, curvas, 426
 —, teorema da curva, 426

K

Kolmogorov, A., 526

L

Lagrange, coeficientes interpolação, 235
 —, interpolação, invariância coeficientes (ex. 6), 659
 —, multiplicadores, 353
 Laplace, equação, 319
 —, Pierre de, 526, 638
 Lebesgue-Stieltjes, integral, 607
 Legendre, equação, 194
 —, polinômio, 28, 197
 —, polinômios, fórmula Rodrigues, 199
 Leis dos grandes números, 634
 Leverrier, Jean Joseph, 643
 Límites campos escalares e vectoriais, 278
 — iterados, 282
 Lindeberg, condição, 639
 —, J. W., 638
 Linhas equipotenciais, 376
 — fluxo, 299
 — integração, 363
 Lipschitz, condição, 257
 Lyapunov, A., 638

M

Matriz ampliada, 69

Matriz associada matriz, 138
 — coeficientes incógnitas sistema, 66
 — coluna, 51, 58
 — complementos algébricos, 103
 — exponencial, 221
 — hessiana, 346
 — —, valores próprios, natureza ponto estacionaridade, 348
 — identidade, 60
 — linha, 58
 — quadrada, 58
 — singular, 74
 — transposta, determinante, 102
 — unidade, 60
 — wronskiana, 107, 180
 — — n soluções independentes equação linear homogênea, não singularidade, 182

Matrizes diagonais, representação transformações lineares, 109
 —, espaços lineares, 58
 — hermiticas e hemi-hermiticas, 138
 — mesma transformação linear, 123
 —, multiplicação, 61
 —, normas, 218
 — ortogonais, 139
 — quadradas, inversas, 73
 —, regra produto exponenciais, 224
 — semelhantes, 123
 —, séries, 218
 — unitárias, 139
 Máximo absoluto, 342
 — relativo, 342, 343, 346
 Máximos, 341
 Medi finitamente aditiva, 528
 Método Frobenius, 204
 — Picard, 250
 Mínimo relativo, 342, 344, 346
 Mínimos, 341
 Möbius, F., 509
 —, superfície, 509
 Moivre, Abraham de, 525
 Momento inércia, 482
 Mudança variáveis, demonstração fórmula caso geral, 453
 — — — — — particular, 450
 — — integral n -múltiplo, 457
 Multiplicação por um escalar fixo, 36
 Multiplicadores Lagrange, 353

N

Napier, 644
 Neptuno, 643

Newton, fórmula interpolação, 662
 Norma máxima, 648
 — quadrática, 647
 Nulidade e ordem, 38
 Número relativo a norma maximal, problema, 699
 Números Stirling primeira espécie (ex. 4), 664

O

Operador derivação, 36
 — diferenças sucessivas, 664
 — integração, 37
 — laplaciano n -dimensional (ex. 3), 330
 Operadores coeficientes constantes, álgebra, 168
 — contração, 263
 — hermiticos e hemi-hermiticos, representação matricial, 137
 Ortogonalidade espaço euclidiano, 20
 — vectores próprios, 133
 Ortogonalização Grám-Schmidt, 25
 Oscilação, 361

P

Pappus de Alexandria, 422
 —, teorema, 479
 —, teoremas, 422
 Partições retângulos, 398
 Pascal, Blaise, 525
 Picard, método, 250
 Plano tangente superfície nível, 300
 Planos tangentes campo escalar, 298
 Poisson, distribuição, 588
 Polinómio Legendre, 28
 — Taylor, 645
 Polinómios característicos, 116
 — factoriais, 666
 — interpoladores, 652
 — Legendre, 197
 — —, fórmula Rodrigues, 199
 — n variáveis, continuidade, 280
 — Tchebycheff, 670
 — —, propriedade mínimo, 672
 — trigonométricos, 33
 Ponto crítico, 342
 — estacionaridade, 342
 — —, natureza pelos valores próprios matriz hessiana, 348
 — exterior, 276
 — fixo, aplicações teorema, 266

Ponto fixo operadores contração, teorema, 264
 — fronteiro, 276
 — interior, 274
 — sela, 342, 344
 Pontos fixos operadores, 261
 — interpolação igualmente espaçados, 655, 664
 — sela, 341
 Potencial conjunto convexo, construção, 387
 — newtoniano, 377
 — num retângulo aberto, construção, 384
 Princípio conservação energia mecânica, 377
 — cosseno áreas, 477
 Probabilidade condicionada, 545
 — conjuntos fundamentais, definição, 567
 — — — não numeráveis, definição, 573
 —, definição para conjuntos fundamentais finitos, 530
 — elementar, 525
 —, exemplos resolvidos, 535
 —, experiências compostas, 551
 —, tiragens com reposição, 542
 —, — sem reposição, 542
 Probabilidades, cálculo, 573
 —, —, teorema limite central, 637
 Produto exponenciais matrizes, regra, 224
 — interno, 16, 36
 — — transformações lineares, 129
 — vectorial fundamental definindo uma normal à superfície, 474
 Projeção sobre subespaço, 36
 Projeções, 30
 Ptolomeu, Claudio, 644
 Putzer, método cálculo e^A , 231

R

Região integração, 397
 Regra Simpson integração, 680
 Representação matricial forma diagonal, 54
 — — transformações lineares, 51
 Resultado (elemento conjunto fundamental), 533
 Riemann-Stieltjes, integral, 607
 Rotacional campo vectorial, 493
 —, divergência, 498
 — gradiente, 497
 — nulo, campo vectorial, 498
 —, propriedades, 496
 — um rotacional, 498

S

- Semi-norma interpolação, 647
- Taylor, 647
- Simetria hermitica, 17
- Simpson, regra integração, 680
- Sistema linear geral, 244
- Sistemas equações diferenciais, introdução, 215
- lineares homogêneos, resolução séries potências, 248
- —, teoremas existência e unicidade, 225
- — não homogêneas com coeficientes constantes, 239
- não lineares primeira ordem, teorema existência e unicidade, demonstração, 257
- Stirling, fórmula, 692
- , Jacob, 644
- , números primeira espécie (ex. 4), 667
- Stokes, G. G., 490
- , teorema, 490
- , —, extensões, 507
- Subespaços espaço linear, 9
- Superfície, equação vectorial, 468
- Möbius, 509
- nível, 299
- , producto vectorial fundamental, 471
- , representação paramétrica, 467
- Superfícies equipotenciais, 376

T

- Taylor, Brook, 644
- , fórmula segunda ordem campos escalares, 346, 347
- , polinómio, 645
- primeira ordem, fórmula, 290
- , semi-norma, 647
- Tchebycheff, desigualdade, 632
- , P. L., 633
- , polinómios, 670
- , —, propriedade mínimo, 672
- Teorema Cayley-Hamilton, 228
- continuidade uniforme campos escalares contínuos, 361
- da curva de Jordan, 426
- divergência, 490, 511
- —, aplicações, 515
- Euler funções homogêneas, 324
- existência, demonstração aproximações sucessivas, 250
- Gauss, 511

Teorema Green, aplicações, 429

- — no plano, 425
- — regiões multiplamente conexas, 435
- — planas limitadas, 427
- limitação campos escalares contínuos, 358
- limite central cálculo probabilidades, 637
- Pappus, 422, 479
- ponto fixo, aplicações, 266
- — operadores contracção, 264
- Stokes, 490
- —, extensões, 507
- valor extremo campos escalares contínuos, 358
- valores extremos campos escalares, 360
- Teoria probabilidades, terminologia, 533
- Trabalho como integral linha, 369
- Traço matriz, 120
- Transformação identidade, 36, 37
- zero, 36, 37
- Transformações hemi-hermíticas, 130
- hermiticas, 130
- lineares, 35
- — biunivocas, 46
- — com valores determinados, 50
- —, exemplo, 447
- —, operações algébricas, 41
- — por matrizes diagonais, 109
- —, representação matricial, 51
- unitárias, 155

U

- Universo, 530
- Urano, 643

V

- Valores próprios, 109
- — dimensão finita, cálculo, 117
- — operadores espaços euclidianos, 129
- — hermiticos e hemi-hermiticos, 132
- — transformação simétrica, 152
- — —, propriedades extremas, 154
- — transformações lineares, 129
- Variância, 625
- Variáveis aleatórias, 575
- — bidimensionais, distribuição, 610
- — (duas) independentes, máximo e mínimo, 621
- —, soma e diferença, 620
- Variável estocástica, 576

Veblen, Oswald, 426

Vector coluna, 51

Vectores próprios, 109

— — dimensão finita, cálculo, 117

— —, independência linear, 113

— — operadores hermiticos e hemi-hermiticos, 132

— — — — —, conjunto ortonormal, 135

— —, ortogonalidade, 133

W

Wallis, desigualdades, 692

—, John, 692

Wronski, J. M. H., 180

Y

Young, W. H., 290